Bridging Theory and Data

Uncertainty-Aware Analyses for the LHC and Beyond

Dissertation

Nina Marie Elmer

Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences
of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Nina Marie Elmer

born in Bünde, Germany

Oral examination: 11.07.2025

Bridging Theory and Data

Uncertainty-Aware Analyses for the LHC and Beyond

Referees: Prof. Dr. Tilman Plehn
apl. Prof. Dr. Monica Dunford

Abstract

Uncertainties are crucial in particle physics, affecting experimental data and theoretical predictions. This thesis investigates the impact of uncertainties on global analyses and the estimation of uncertainties using machine learning architectures. In the first part of this thesis, we perform global analyses using effective field theory approaches. We start with the Standard Model effective field theory in the top, Higgs, and electroweak sectors, including public experimental likelihoods. In particular, we focus on the role of theory uncertainties and their interplay with correlations. Next, we perform the first global electric dipole moment analysis constraining model parameters from the hadronic- and weak-scale Lagrangians while exploring the impact of theory uncertainties on the parameter constraints. The second part discusses machine-learning methods that have become increasingly important with the growing data from future LHC runs. Thus, we study the calibration of systematic and statistical uncertainties and the precision and reliability of machine learning architectures for amplitude surrogate models. We compare Bayesian neural networks and repulsive ensembles as uncertainty estimators regarding their precision and use Kolmogorov-Arnold networks to explore the impact of activation functions.
Overall, this work emphasizes the importance of reducing theory uncertainties and paves new ways of uncertainty estimation using machine learning models in particle physics and global analyses.

Zusammenfassung

Unsicherheiten sind in der Teilchenphysik von entscheidender Bedeutung für experimentelle Daten und theoretische Vorhersagen. In dieser Arbeit werden die Auswirkungen von Unsicherheiten in verschiedenen globalen Analysen und auch die Abschätzung von Unsicherheiten im Bereich des maschinellen Lernens untersucht. Zunächst werden globale Analysen mithilfe effektiver Feldtheorien durchgeführt. Als erstes wird die effektive Feldtheorie des Standardmodells im Top-, Higgs- und elektroschwachen Sektor unter Verwendung von öffentlichen, experimentellen Likelihood-Funktionen eingeschränkt. Dabei liegt ein Fokus auf der Rolle der Theorieunsicherheiten und deren Zusammenspiel. Als nächstes wird die erste globale Analyse mit elektrischen Dipolmomenten durchgeführt, die Modellparameter aus den hadronischen und schwachen Lagrangedichten einschränkt, wobei wir die Auswirkungen von Theorieunsicherheiten auf die Parameterbereiche untersuchen. Der zweite Teil der Arbeit befasst sich mit Methoden des maschinellen Lernens, die angesichts der wachsenden Datenmengen am LHC immer wichtiger werden. Zu diesem Zwecke werden die Kalibrierung von systematischen und statistischen Unsicherheiten, sowie die Präzision und Zuverlässigkeit von Architekturen des maschinellen Lernens für Amplitudensurrogatmodelle untersucht. Dazu werden Bayes'sche neuronale Netze und repulsive Ensembles als Unsicherheitsschätzer hinsichtlich ihrer Präzision verglichen und Kolmogorov-Arnold-Netze verwendet, um die Auswirkungen verschiedener Aktivierungsfunktionen zu untersuchen.
Insgesamt unterstreicht diese Arbeit die Bedeutung der Reduzierung theoretischer Unsicherheiten und zeigt neue Wege der Unsicherheitsabschätzung mit Hilfe von Modellen des maschinellen Lernens in der Teilchenphysik und bei globalen Analysen auf.

Contents

Preface

The research presented in this thesis was conducted at the Institute for Theoretical Physics at Heidelberg University from November 2022 to April 2025. The contents of Chapters 4, 5, and 7 are based on work in collaboration with other researchers and have been previously published as

[1] Nina Elmer, Maeve Madigan, Tilman Plehn, Nikita Schmal,
Staying on Top of SMEFT-Likelihood Analyses,
SciPost Phys. 18 (2025) 108, arXiv:2312.12502 [hep-ph];

[2] Skyler Degenkolb, Nina Elmer, Tanmoy Modak, Margarete Mühlleitner, Tilman Plehn,
A Global View of the EDM Landscape,
submitted to SciPost, arXiv:2403.02052 [hep-ph];

[3] Henning Bahl, Nina Elmer, Luigi Favaro, Manuel Haußmann, Tilman Plehn, Ramon Winterhalder,
Accurate Surrogate Amplitudes with Calibrated Uncertainties,
submitted to SciPost, arXiv:2412.12069 [hep-ph]

Finally, the author is involved in ongoing projects that have not been ready for publication at the time of writing this thesis.

Chapter 1

Introduction

The Standard Model (SM) is the mathematical foundation of particle physics, describing fundamental particles and their interactions. It has been completed by discovering the Higgs boson at the Large Hadron Collider (LHC) in 2012 [4,5]. Since then, the predictions of the SM have become more and more precise, but some questions remain open. Such unanswered questions address the nature of dark matter, the origin of baryon asymmetry, or the neutrino mass. Experimental evidence in astroparticle physics and cosmology indicates these open questions that currently cannot be answered by the SM. However, there are no traces of such new physics or physics beyond the Standard Model (BSM) in the LHC data so far [6–15]. One existing idea to solve these questions is to extend the SM to higher orders in perturbation theory, describing it by an effective field theory (EFT). This EFT approach of the SM, the Standard Model effective field theory (SMEFT), assumes new physics hidden at higher energy scales, which the LHC does not reach. Thus, these new particles are produced off-shell and cannot be directly probed with LHC searches. Building on the SM, SMEFT describes a model-independent approach for new physics searches. However, in its low-energy limit, it must retain the SM.

Given the sheer amount of data produced by the LHC in recent years, a single-parameter analysis, including only single measurements, cannot include all the effects and correlations between different measurements and parameters. When performing a global analysis, all effects hidden in the data and model, such as possible interplays and connections between different SMEFT parameters, must be included. Thus, we have to perform a global analysis in the SMEFT framework, including data from Higgs, top processes, and electroweak precision observables (EWPOs). Additionally, we incorporate measurements with publicly available likelihoods from ATLAS top production processes and determine their impact on the overall analysis. Including these experimental likelihoods simplifies the treatment of correlations and uncertainties in a global analysis by providing detailed information about different uncertainty categories and their correlations. In the global SMEFT analysis, uncertainties within the ATLAS and CMS experiments are treated as correlated, while only the Luminosity is correlated between both experiments. Furthermore, we consider systematic, statistical, and theory uncertainties and investigate the impact of theory uncertainties on the constraints of the SMEFT parameters.

SMEFT is an excellent example of an EFT in the high-energy limit. However, the EFT approach can also address problems within a low-energy limit. An example of an EFT in the low-energy limit is a weak-scale EFT describing processes at the GeV scale. This weak-scale EFT can describe electric dipole moment (EDM) measurements and link them to SMEFT dimension six operators. As mentioned before, the SM fails to explain the baryon asymmetry in the universe. The Sakharov conditions [16] do not solve

that question but provide structures that are required to answer this question. These requirements are: (i) Charge (C) and Charge-Parity (CP) violation, (ii) baryon number violation, and (iii) a deviation from the thermal equilibrium. The first requirement of C- and CP-violation can be read off from fundamental Lagrangians, like the SM Lagrangian. However, with the neutron EDM measurement agreeing with zero, the CP violation arising from the quantum chromodynamic (QCD) phase in the SM is too small to explain the observed baryon asymmetry [17–22]. A new approach is needed to explain this baryon asymmetry and the resulting CP violation. If the measured EDMs exclude zero, they are a possible explanation for this CP violation. Thus, they are among the best hints for CP violation and baryon asymmetry. EDM measurements include a wide range of different atoms and molecules [23–26], but up until now, they have only been evaluated using single parameters and measurement estimates. We extend this evaluation to a global approach using an EFT based on weak-scale [24, 27–32] and hadronic Lagrangians [23, 26, 27, 33]. Thus, EDMs can be introduced in a global analysis, similar to the SMEFT framework, to simultaneously evaluate several measurements and parameters.

With the entering of the High Luminosity (HL) phase of the LHC, a large amount of data will be generated, and thus they probe the predictions with more precision [34]. Previously used tools for data evaluation, like Monte Carlo Markov Chains (MCMCs), will face computational problems regarding speed and data storage. At this point, machine learning (ML) techniques need to enter the physicist's toolbox as a more efficient tool. These techniques can consider every aspect of the provided data. In particle physics, we include many experimental measurements and precise simulations of theoretical predictions for these measurements from first principles [35], which leads to a large data set size. Hence, ML applications can be found in various aspects of LHC physics. Some examples are generative networks for event generation and simulation [36–44] and surrogates for amplitude regression tasks [45–50]. This thesis focuses on the benefits of ML algorithms applied to amplitude surrogates for a loop-induced gluon fusion process. In this task of amplitude regression, results with controlled accuracy and precision are required, including handling different uncertainty types, such as systematic and statistical uncertainties. These amplitude surrogate models are part of ongoing research [51].

Thus, ML architectures can combine the speed and precision in their predictions with a controlled uncertainty estimation. These uncertainties arising from measurements are important to take into account, as are uncertainties originating from the network architecture. For this purpose, we use Bayesian Neural Networks (BNNs) [52–55], Repulsive Ensembles (REs) [56–58] and a Neural Network (NN) with a heteroscedastic loss. They provide systematic and statistical uncertainties by their sampling nature.

This thesis is organized in the following way: We introduce the basic concepts of EFTs for SMEFT and how EDMs are connected to a low-energy EFT approach in Ch. 2. Next, SFitter as the tool of choice for the global analyses is presented in Ch. 3. We also discuss the advantages of MCMCs. Ch. 4 and Ch. 5 present the results from the global analyses in the SMEFT and EDM sectors. The SMEFT global analysis focuses on implementing likelihood measurements in SFitter and their impact on uncertainties and correlations. Chapter 4 ends with a combined analysis using Higgs, top, and EWPO data. The global analysis performed on EDM measurements in Ch. 5 focuses on the impact of theory uncertainties and different measurements on the global analysis results. We then move on to some introduction to ML algorithms and discuss different network architectures like BNNs and REs in Ch. 6.

Then, Ch. 7 describes the uncertainty estimation using BNNs, REs, and more advanced

architectures on amplitude surrogates. We also focus on the impact of network expressivity and data preprocessing on the intrinsic uncertainties. Finally, we summarize the results and discuss future research directions in Ch. 8.

Chapter 2

Effective Field Theories and SMEFT

2.1 The Effective Field Theory Formalism

This chapter starts with an introduction to the EFT formalism, following the approach of Refs. [59–66]. An EFT is a powerful tool used to describe physical systems in well-defined regions. With the incomplete SM, we can assume that BSM physics occurs at higher orders $\Lambda \gg v$. Where $v = 246.22$ GeV is the cut-off scale for electroweak processes, at which the SM is still valid. Hence, we aim to describe the deviations from the SM in a model-independent and reproducible way. New physics can be discovered through direct searches, like bump-hunts, which led to the discovery of the Higgs boson, and indirect searches, like deviations in the high-energy tails of distributions. These subtle deviations in indirect searches often hint at a discovery before it is actually revealed by a direct detection method. Due to their higher-order expansion, EFTs can cover these subtle deviations in searches for new physics. Thus, they are a versatile and model-independent way to start the search for BSM physics at higher energy scales Λ. First, we explain the general idea and construction of an EFT. Next, in Sec. 2.1.2, we continue to construct SEMFT as an EFT build-up from the SM. We will list all necessary ingredients, such as symmetries, fields, and additional operators used in a global analysis. Additionally, we consider EDMs as low-energy EFT obtained from the weak-scale and hadronic-scale Lagrangians.

2.1.1 Constructing an EFT

As a quantum field theory (QFT) an EFT is built from a set of fields and their corresponding derivatives as building blocks of the effective Lagrangian $\mathscr{L}_{\text{eff}}$. Besides the field content, the EFT is also defined by its underlying symmetries, which are encoded in the Lagrangian. Keeping these two building blocks in mind, we can start by constructing $\mathscr{L}_{\text{eff}}$ from all possible operators $\mathcal{O}_i$, which are invariant under the underlying symmetries, and coefficients c_i:

$$\mathscr{L}_{\text{eff}} = \sum_i c_i \mathcal{O}_i \,. \tag{2.1}$$

Compared to a QFT, which is usually renormalizable and removes every term with dimension $D > 4$, an EFT does not have to be renormalizable. The effective Lagrangian of an EFT shown in Eq. (2.1) contains a sum over an infinite number of operators $\mathcal{O}_i$. This infinite number of operators requires an endless amount of counter-terms, leading to a non-renormalizable theory. However, in Sec. 2.1.2 we show, that EFTs can be truncated

to a finite number of operators resulting in a renormalizable theory. Without trunctation, the infinite number of operators in $\mathscr{L}_{\text{eff}}$ naturally leads to the question of how it can be used to predict BSM physics?

This question can be addressed by introducing the method of power counting. Power-counting motivates the truncation of higher-order terms in the Lagrangian, reducing the number of operators to a finite number and making the theory renormalizable. An EFT only effectively describes new physics at higher energy scales by construction. With this, the effective description is only valid in certain phase space regions. We can define these regions by introducing a cut-off energy Λ and assume that our constructed EFT is only valid for energies $E \ll \Lambda$. By introducing that cut-off scale, we also introduce an energy dependence to the effective Lagrangian, and the EFT becomes an expansion in terms of $\delta = E/\Lambda$. This enables us to truncate Eq. (2.1) at a fixed order regarding the energy ratio δ. This truncation transforms the infinite number of operators into a finite number, resulting in a renormalizable theory. By applying the condition $\delta \ll 1$, we ensure that our phase space region is within the allowed description of the EFT. Additionally, by assuming different mass and energy dimensions for various operators, the significance of certain operators in relation to specific observables may change. Introducing this power counting scheme, we address a specific value for δ to every operator. Using this assigned value of δ, the operators and coefficients in the effective Lagrangians can be ordered in terms of δ, with a mass dimension D_i for the operators $\mathcal{O}_i$. To produce a dimensionless action of our effective Lagrangian $[S(\mathscr{L}_{\text{eff}})] = 0$, we require $[\mathscr{L}_{\text{eff}}] = 4$. This implies a mass dimension of $[c_i] = 4 - D_i$ for the coefficients in the effective Lagrangian. However, the different mass dimensions of the different c_i make it challenging to formulate a consistent and uniform basis for the operators and coefficients in global analyses later on. To make the comparison and interplay between several operators and their coefficients more apparent, we introduce dimensionless Wilson coefficients (WCs) C_i. Therefore, the dimensionality of the coefficients c_i is factored out in powers of the energy scale Λ. This leads to the new formulation of the effective Lagrangian

$$\mathscr{L}_{\text{eff}} = \sum_i \frac{C_i}{\Lambda^{D_i-4}} \mathcal{O}_i \,. \tag{2.2}$$

By construction, every operator that enters an arbitrary process at tree level is accompanied by a factor $\frac{1}{\Lambda^{D_i-4}}$, and thus, the operator has mass dimension $n = D_i - 4$. If multiple operators are inserted, their individual contributions sum up to a total contribution of

$$n = \sum_i (D_i - 4) \,. \tag{2.3}$$

We can now use this power-counting formula for n to determine where to truncate our Lagrangian from Eq. (2.2) to include specific orders n. If we only want to include leading order (LO) terms in the EFT expansion with $n = 0$, only terms with dimension $D \leq 4$ are included. If we want to go to next-to-leading order (NLO) terms with $n = 1$ and $D = 5$ dimensional operators are included, and with next-to-next-to-leading order (NNLO) terms equal to $n = 2$ are included. These terms can either be a quadratic $D = 5$ contribution or a linear $D = 6$ contribution. This new ordering of operators in terms of dimension allows us to easily determine where to truncate our EFT and which operators to consider. The final expression for the effective Lagrangian reads

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{D\leq 4} + \sum_i \frac{C_i^5}{\Lambda} \mathcal{O}^5 + \sum_i \frac{C_i^6}{\Lambda^2} \mathcal{O}^6 + \dots \,. \tag{2.4}$$

The first part of the effective Lagrangian in Eq. (2.4), $\mathscr{L}_{D\leq 4}$, describes the LO terms in δ. These include all the relevant operators in a renormalizable QFT, such as the SM operators. The following terms are the NLO and NNLO operators in terms of δ. These are subleading in n and thus neglected in the QFT. If we move to lower scales of the energy E, they are suppressed and become less important. This shows that the constructed EFT provides a theory that is only valid in well-defined phase-space regions. It describes deviations from SM physics at given energy scales without knowing all the details from higher energy scales. Also, by introducing a new energy scale Λ, we implicitly assume that the particles, which might enter in BSM physiscs, are heavy and thus are integrated out in the low-energy limit $E \ll \Lambda$.

To construct an EFT based on a renormalizable QFT we can choose between a top-down or bottom-up approach. In the top-down approach the EFT is treated as a limit of a well-known, renormalized QFT in a low-energy regime. It starts from the complete high-energy theory which is then matched to a low-energy theory. Therefore, we integrate out the heavy particles, or so-called degrees of freedom (DOF). Next, we let the couplings run to the appropriate scale of the low-energy processes. Thus, the corresponding WCs can be treated as a function of the high-energy theory parameters. One example of a top-down approach is the weak effective field theory (WET), in which we reduce the full SM to the weak scale and integrate out the top quark, W, Z, and Higgs boson.

The opposite of the top-down approach is the bottom-up approach, in which we take an existing theory as a starting point and build an EFT based on the particle content of this low-energy theory and (non-accidental) symmetries. The constructed EFT is then agnostic about new physics outside its validity. However, we have to find all the independent operators that can be built from the assumed low-energy particle content and underlying symmetries. In principles there are infinitely many operators, leading to a non-renormalizable theory. To make the constructed EFT renormalizable again, we can use the previously introduced concept of power counting by ordering all new operators in expansion as a separation of scale. Then, we can define an energy cut-off in δ and only keep the necessary terms for our theory. The resulting WCs and operators parameterize the new physics in terms of low-energy degrees of freedom and symmetries that are already known. This provides an effective theory that is UV model-independent, and by truncation at a particular order, it is also renormalizable.

2.1.2 SMEFT

One example of an effective theory, built by using the bottom-up approach, is SMEFT. The SM is an excellent description up to the EW scale with a vacuum expectation value (vev) around $v \sim 246$ GeV. BSM physics might not be visible at such low scales because it may consist of new, heavy particles at a scale $\Lambda \gg v$. Thus, they can be treated as integrated out of the UV complete theory. This results in operators of dimension $D \geq 5$, consisting of interactions using the SM degrees of freedom. We will use the SMEFT framework to break down and parameterize the effects of these high-mass fields on energy scales accessible by the LHC [59, 62, 65, 67–70]. Therefore, we start by writing down the symmetries and particle content of the SM to construct the SMEFT framework. The most critical SM symmetry is the gauge symmetry group

$$SU(3)_c \times SU(2)_L \times U(1)_Y \ , \tag{2.5}$$

Field	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
$q_i = \{u_{L,i}, d_{L,i}\}$	3	2	1/6
$u_{R,i}$	3	1	2/3
$d_{R,i}$	3	1	$-1/3$
$l_j = \{\nu_{L,j}, e_{L,j}\}$	1	2	$-1/2$
$e_{R,j}$	1	1	-1
ϕ	1	2	1/2
G	8	1	0
W	1	3	0
B	1	1	0

Table 2.1: Conventions for the SM field content and their corresponding quantum numbers. The indices $i, j \in \{1, 2, 3\}$ denote the three different fermion families of leptons and quarks and L, R denote left- and right-handed fermions.

where the first term describes the strong interactions and the remaining ones account for the electric and weak forces. The $SU(2)_L \times U(1)_Y$ combination is realized linearly, and the Higgs is represented as a doublet in the $SU(3)_c$ group. Another global symmetry of the SM is the Charge-Parity-Time (CPT) symmetry and a spacetime symmetry, realized by a Poincaré $ISO(3,1)$ group. As for the particle content, displayed in Tab. 2.1, we assume three generations of fermions, where ϕ denotes the content of the Higgs doublet and G, W and B representing the gauge fields of the $SU(3)_c$, $SU(2)_L$ and the $U(1)_Y$ respectively. The conventions for the three fermion generations are shown in Table 2.1.

With the field and particle content of the SM, we can construct the SMEFT Lagrangian in the following way,

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_{D>4} \sum_{i} \frac{C_i}{\Lambda^{D-4}} \mathcal{O}_i^D \,, \tag{2.6}$$

where $\mathscr{L}_{\text{SM}}$ denotes the SM Lagrangian, and the other term denotes the higher-order expansion.

The physical SM Lagrangian reads as follows

$$\begin{aligned}
\mathscr{L}_{\text{SM}} &= \mathscr{L}_{\text{fermion}} + \mathscr{L}_{\text{Yukawa}} + \mathscr{L}_{\text{Gauge}} + \mathscr{L}_{\text{Higgs}} \\
\mathscr{L}_{\text{fermion}} &= i \sum_j \left(\bar{l}_{L,j} \not{D} l_{L,j} + \bar{e}_{R,j} \not{D} e_{R,j} + \bar{q}_{L,j} \not{D} q_{L,j} + \bar{d}_{R,j} \not{D} d_{R,j} + \bar{u}_{R,j} \not{D} u_{R,j} \right) \\
\mathscr{L}_{\text{Yukawa}} &= -\sum_j \left(\bar{l}_{L,j} Y_l \phi e_{R,j} + \bar{q}_{L,j} Y_d \phi d_{R,j} + \bar{q}_{L,j} Y_u \tilde{\phi} u_{R,j} + \text{h. c.} \right) \\
\mathscr{L}_{\text{Gauge}} &= -\frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} G^A_{\mu\nu} G^{A\mu\nu} \\
\mathscr{L}_{\text{Higgs}} &= (D_\mu \phi)^\dagger D^\mu \phi + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{2} \left(\phi^\dagger \phi \right)^2 .
\end{aligned} \tag{2.7}$$

The sum over j in the $\mathscr{L}_{\text{fermion}}$ and $\mathscr{L}_{\text{Yukawa}}$ indicate the sum over the three fermion generations. At the same time, $Y_{l,d,u}$ are the different Yukawa matrices in the generation space, and $\tilde{\phi} = i\sigma^2 \phi^\star$, with the Pauli matrices σ. In $\mathscr{L}_{\text{Gauge}}$, the upper case indices span the color space and the lower case ones the isospin space, with G^A representing the eight

gluons and (W^a, B) the four electroweak gauge bosons.

Next, we discuss the second part of the SMEFT Lagrangian in Eq. (2.6), consisting of operators with dimensions five, six, and higher-order operators. There, we have to consider all operators that can be constructed from the Lorentz invariant operators of the SM field content concerning the SM gauge symmetries.

The Dimension 5 Operator

At dimension five, only one new operator contributes to the SMEFT. This operator is called the Weinberg operator [71] and is given by

$$\mathcal{O}_{ij}^5 = \left(\bar{l}_i^c \tilde{\phi}^\star\right)\left(\tilde{\phi}^\dagger l_j\right) + \text{h.c.}\,, \tag{2.8}$$

where $\tilde{\phi}$ is the same as in the Yukawa Lagrangian in Eq. (2.8) and l^c denotes the charge-conjugated leptons with i, j as indices for the lepton generation. This operator violates the lepton number conservation, but it is allowed in SMEFT because the lepton number conservation is only an accidental symmetry of the SM and thus does not have to be preserved. Also, the Weinberg operator generates the mass and mixing terms for the neutrinos. After the electroweak symmetry breaking (EWSB), the $\mathcal{O}^5$ is reduced to

$$\mathcal{O}^5 \sim \frac{v^2}{2}\bar{\nu}_L^c \nu_L\,, \tag{2.9}$$

which indicates a Majorana mass term. The detection of neutrino oscillations [72] first provided the existence of this phenomenon of a non-zero neutrino mass difference Δm^2.

Dimension 6 Operators

When going from dimension four to dimension five, we only have to take one additional operator into account, but the number of operators gets larger as soon as we also take dimension six operators into account. At dimension six, 63 new operators are entering. Of those newly entering operators, 59 are baryon number conserving, and additional four operators are baryon number violating. This is without considering all the different flavor structures, CP violating operators and hermitian conjugates. Considering these operators, the number of dimension six operators grows to 2499 [73]. Different conventions, also known as bases, can be used to build these operators. The two most used bases are the Warsaw basis [70] and the HISZ basis [74]. Both bases reduce the number of operators by applying different equations of motions and integration by parts as tool. It is also possible to transform one basis into another by applying a basis rotation. However, this rotation may change the process class and category of the operator since an operator that consists purely of the Higgs doublet on an HISZ basis might get an additional four-fermion interaction when rotated into the Warsaw basis. Additionally, a complete list of all 2499 operators is provided in Ref. [73]. The SMEFT global analysis, discussed in Chapter 4, takes 43 operators into account from the top, Higgs, and electroweak sectors and the invisible decay of the Higgs. The following list of operators is taken from Refs. [1] and [75]. Considering more operators would exceed the number of available measurements to obtain constraints for all these operators. With this, we truncate the

effective Lagrangian to

$$\mathscr{L}_{\text{eff}} = \sum_j \left(\frac{C_j}{\Lambda^2} \, {}^{\ddagger}\mathcal{O}_j + \text{h.c.} \right) + \sum_k \frac{C_k}{\Lambda^2} \, \mathcal{O}_k \,, \tag{2.10}$$

where ${}^{\ddagger}\mathcal{O}_k$ denotes non-hermitian operators. We neglect the Weinberg operator at dimension five and operators exceeding dimension 6. This choice is motivated by the Λ-suppression of the higher order terms, which can be assumed to directly translate to suppressing their effects on LHC observables. We only consider CP-conserving operators and assume separate U(2) symmetries in the first and second quark generations [76, 77]

$$\begin{aligned} q_i &= (u_L^i, d_L^i) \qquad & u_i &= u_R^i,\, d_i = d_R^i \quad \text{for} \quad i = 1,2 \\ Q &= (t_L, b_L) \qquad & t &= t_R,\, b = b_R \,, \end{aligned} \tag{2.11}$$

and a lepton U(3) symmetry. In the quark sector, all masses except the one of the top quark are assumed to be zero.

With these assumptions, we are left with 22 different operators in the top sector. These operators can be split up into three groups. First, we have eight operators that encounter a chiral LL or RR structure when interacting with fermions

$$\begin{aligned} \mathcal{O}_{Qq}^{1,8} &= (\bar{Q}\gamma_\mu T^A Q)\,(\bar{q}_i \gamma^\mu T^A q_i) & \mathcal{O}_{Qq}^{1,1} &= (\bar{Q}\gamma_\mu Q)\,(\bar{q}_i \gamma^\mu q_i) \\ \mathcal{O}_{Qq}^{3,8} &= (\bar{Q}\gamma_\mu T^A \tau^I Q)\,(\bar{q}_i \gamma^\mu T^A \tau^I q_i) & \mathcal{O}_{Qq}^{3,1} &= (\bar{Q}\gamma_\mu \tau^I Q)\,(\bar{q}_i \gamma^\mu \tau^I q_i) \\ \mathcal{O}_{tu}^{8} &= (\bar{t}\gamma_\mu T^A t)\,(\bar{u}_i \gamma^\mu T^A u_i) & \mathcal{O}_{tu}^{1} &= (\bar{t}\gamma_\mu t)\,(\bar{u}_i \gamma^\mu u_i) \\ \mathcal{O}_{td}^{8} &= (\bar{t}\gamma^\mu T^A t)\,(\bar{d}_i \gamma_\mu T^A d_i) & \mathcal{O}_{td}^{1} &= (\bar{t}\gamma^\mu t)\,(\bar{d}_i \gamma_\mu d_i) \,. \end{aligned} \tag{2.12}$$

Next, six operators contain a LR or RL chiral structure in current-current interactions

$$\begin{aligned} \mathcal{O}_{Qu}^{8} &= (\bar{Q}\gamma^\mu T^A Q)\,(\bar{u}_i \gamma_\mu T^A u_i) & \mathcal{O}_{Qu}^{1} &= (\bar{Q}\gamma^\mu Q)\,(\bar{u}_i \gamma_\mu u_i) \\ \mathcal{O}_{Qd}^{8} &= (\bar{Q}\gamma^\mu T^A Q)\,(\bar{d}_i \gamma_\mu T^A d_i) & \mathcal{O}_{Qd}^{1} &= (\bar{Q}\gamma^\mu Q)\,(\bar{d}_i \gamma_\mu d_i) \\ \mathcal{O}_{tq}^{8} &= (\bar{q}_i \gamma^\mu T^A q_i)\,(\bar{t}\gamma_\mu T^A t) & \mathcal{O}_{tq}^{1} &= (\bar{q}_i \gamma^\mu q_i)\,(\bar{t}\gamma_\mu t) \,. \end{aligned} \tag{2.13}$$

Finally, there is an additional set of eight operators which couple two heavy quarks to gauge bosons [78],

$$\begin{aligned} \mathcal{O}_{\phi Q}^{1} &= (\phi^\dagger \, i \overleftrightarrow{D_\mu} \, \phi)\,(\bar{Q}\gamma^\mu Q) & {}^{\ddagger}\mathcal{O}_{tB} &= (\bar{Q}\sigma^{\mu\nu} t)\,\widetilde{\phi}\, B_{\mu\nu} \\ \mathcal{O}_{\phi Q}^{3} &= (\phi^\dagger \, i \overleftrightarrow{D_\mu^I} \, \phi)\,(\bar{Q}\gamma^\mu \tau^I Q) & {}^{\ddagger}\mathcal{O}_{tW} &= (\bar{Q}\sigma^{\mu\nu} t)\,\tau^I \widetilde{\phi}\, W_{\mu\nu}^I \\ \mathcal{O}_{\phi t} &= (\phi^\dagger \, i \overleftrightarrow{D_\mu} \, \phi)\,(\bar{t}\gamma^\mu t) & {}^{\ddagger}\mathcal{O}_{bW} &= (\bar{Q}\sigma^{\mu\nu} b)\,\tau^I \phi\, W_{\mu\nu}^I \\ {}^{\ddagger}\mathcal{O}_{\phi tb} &= (\widetilde{\phi}^\dagger i D_\mu \phi)\,(\bar{t}\gamma^\mu b) & {}^{\ddagger}\mathcal{O}_{tG} &= (\bar{Q}\sigma^{\mu\nu} T^A t)\,\widetilde{\phi}\, G_{\mu\nu}^A \,. \end{aligned} \tag{2.14}$$

The relation of these operators with the Warsaw basis [70] is worked out in the appendix of Ref. [79]. The interactions with the physical states are given by the gauge structure of the electroweak SM, so we use the combinations

$$C_{\phi Q}^{\pm} = C_{\phi Q}^{1} \pm C_{\phi Q}^{3} \qquad \text{and} \qquad C_{tZ} = c_w C_{tW} - s_w C_{tB} \,. \tag{2.15}$$

This way, $C_{\phi Q}^{-}$ and C_{tZ} characterize a $t\bar{t}Z$ interaction, C_{tW} a tbW interaction, and $C_{\phi Q}^{3}$ both tbW and $b\bar{b}Z$ interactions. These are all operators we take into account for the

top sector. To include the Higgs and electroweak sector in the global analysis, we have to extend the list of operators by additional 20 SMEFT parameters and the branching ratio of the Higgs to dark matter agents, BR_{inv}, as a free parameter. The operators are chosen based on Ref [75], rotated from the HISZ to the Warsaw basis to match them with the operators from the top sector. The operators constrained by di-boson, Higgs, and electroweak data can be divided into several groups. First, seven operators are purely constrained by bosonic interactions,

$$\begin{aligned} \mathcal{O}_{\phi G} &= \phi^\dagger \phi G^A_{\mu\nu} G^{A\mu\nu} & \mathcal{O}_W &= \varepsilon^{IJK} W^{I\nu}_\mu W^{J\rho}_\nu W^{K\mu}_\rho \\ \mathcal{O}_{\phi\Box} &= (\phi^\dagger\phi)\Box(\phi^\dagger\phi) & \mathcal{O}_{\phi D} &= (\phi^\dagger D^\mu \phi)^*(\phi^\dagger D^\mu \phi) \\ \mathcal{O}_{\phi B} &= \phi^\dagger \phi B_{\mu\nu} B^{\mu\nu} & \mathcal{O}_{\phi W} &= \phi^\dagger \phi W^I_{\mu\nu} W^{I\mu\nu} \\ \mathcal{O}_{\phi WB} &= \phi^\dagger \tau^I \phi W^I_{\mu\nu} B^{\mu\nu}\,. \end{aligned} \tag{2.16}$$

Next, four single-current operators are modifying the Yukawa couplings,

$$\begin{aligned} \mathcal{O}_{e\phi,22} &= (\phi^\dagger\phi)(\bar{l}_2 \mu \phi) & \mathcal{O}_{e\phi,33} &= (\phi^\dagger\phi)(\bar{l}_3 \tau \phi) \\ \mathcal{O}_{u\phi,33} &= (\phi^\dagger\phi)(\bar{Q}_3 t \phi) & \mathcal{O}_{d\phi,33} &= (\phi^\dagger\phi)(\bar{Q}_3 b \phi)\,, \end{aligned} \tag{2.17}$$

and eight operators that contribute to the electroweak gauge boson coupling to fermions,

$$\begin{aligned} \mathcal{O}_{\phi e} &= (\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{e}\gamma^\mu e) & \mathcal{O}_{\phi b} &= (\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{b}\gamma^\mu b) \\ \mathcal{O}_{\phi d} &= \sum_{i=1}^{2}(\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{d}_i\gamma^\mu d_i) & \mathcal{O}_{\phi u} &= \sum_{i=1}^{2}(\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{u}_i\gamma^\mu u_i) \\ \mathcal{O}^{(1)}_{\phi q} &= \sum_{i=1}^{2}(\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{q}_i\tau^I\gamma^\mu q_i) & \mathcal{O}^{(3)}_{\phi q} &= \sum_{i=1}^{2}(\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{q}_i\gamma^\mu q_i) \\ \mathcal{O}^{(1)}_{\phi l} &= (\phi^\dagger i\overleftrightarrow{D_\mu}\phi)(\bar{l}\gamma^\mu l) & \mathcal{O}^{(3)}_{\phi l} &= (\phi^\dagger i\overleftrightarrow{D}^I_\mu\phi)(\bar{l}\tau^I\gamma^\mu l)\,. \end{aligned} \tag{2.18}$$

Lastly, we also consider the four-lepton operator

$$\mathcal{O}_{ll} = (\bar{l}\gamma_\mu l)(\bar{l}\gamma^\mu l)\,. \tag{2.19}$$

The global SMEFT analysis performed in Ch. 4 includes all these 43 operators.

Yet, one might wonder why we stop at dimension six and do not consider dimension eight operators. Since we cannot measure the WCs and operator contributions directly, we have to extract them from measured physical observables, like cross-sections (total and differentiable) and signal strength measurements. In these physical measurements, the SMEFT parameters appear in linear and quadratic terms, with a respective suppression in terms of Λ^{-2} and Λ^{-4}. These quadratic contributions are entering with the same order as linear dimension eight terms would. So, there are some choices to make: Do we include linear and quadratic dimension six terms, and if we include quadratic contributions, should we then also include linear dimension eight terms? First, including linear and quadratic contributions makes the process of deriving SMEFT predictions in the global analysis more complex and makes the results harder to interpret. However, we want to include both whenever possible since there are some good reasons for doing so. Including both contributions ensures a positive prediction of the physical observable in the SMEFT. If we only include linear terms for some WC combinations, the predictions can get negative, leading to an unphysical result. These unphysical predictions are hard to handle and constrain in global analyses, and up until now, there has been no good way to control

the numerical responses to such negative predictions. Another argument for including the quadratic contribution is the introduction of interference between different WCs. This interference helps prevent some flat directions in the global analysis, which can occur when the data is best described by an interplay of different WCs [79]. Some might argue that stopping after quadratic dimension six contributions is not well motivated because they enter with the same energy suppression of Λ^{-4} as linear dimension eight terms would enter. From a mathematical point of view, this totally makes sense, but the linear dimension eight terms encounter the same problem as the linear dimension six terms, as they can shift the prediction towards negative values, making them unphysical. With dimension eight a large number of new operators would enter in the Lagrangian, making it harder to constrain all operators [80]. Additionally, the present dimension six operators already test a large variety of possible interactions. These are the reasons why we decided to truncate the SMEFT Lagrangian in the global analysis after the quadratic dimension six contributions.

Considering different process classes

As mentioned earlier, 2499 new operators enter the SMEFT when moving to the dimension six expansion. However, we restrict the global analysis to only 43 operators. This restriction is also motivated by the fact that including all 2499 is numerically challenging. Additionally, there are more possible operators than processes which can constrain them at the LHC. Furthermore, the different operators can be assigned to different process classes, which they constrain.

By performing a global analysis, we want to constrain as many different classes of processes at the LHC as possible with a sufficient number of operators and corresponding data. Also, many operators contribute to more than one class, enabling us to cover correlations and interplays between different processes, which is extremely valuable in a global analysis. These process classes include EWPOs, Higgs, and top data. With more subcategories like $t\bar{t}$ or single-t production in the top, or di-boson processes in the Higgs sector. Table 2.2 shows the WCs used in the SMEFT global analysis and what process classes they constrain. Noticeably, most of the operators constrain more than one class, which enhances the cross-talk between several coefficients and measurements.

Based on the choice of WCs included in the global analysis, we constrain all possible process classes and ensure that the global analysis also includes mixing between different classes. In the end, we do not include all operators and orders, but the motivated choice of truncating the Lagrangian beyond dimension six and the careful selection of the 43 WCs ensure that the global analysis performed can provide everything you would expect from a global analysis. Namely, it covers correlations between different operators and processes, covers most of the different processes, and includes a variety of data, like EWPO, Higgs production and decay, top production and decay, and several cross-section, rate, and signal strength measurements.

After discussing the SMEFT as an approach for an EFT at higher energy scales, we can also describe low-energy phenomena in terms of an EFT and effective Lagrangians using the top-down approach. Therefore, we have a look at EDMs and how they can be treated as low-energy EFT.

Operator	Top	Higgs	EW	Operator	Top	Higgs	EW
$\mathcal{O}_{Q,q}^{1,8}$	✓	✓		$\mathcal{O}_{\phi G}$		✓	
$\mathcal{O}_{Q,q}^{1,1}$	✓	✓		$\mathcal{O}_{W}$			✓
$\mathcal{O}_{Q,q}^{3,8}$	✓	✓		$\mathcal{O}_{\phi\Box}$		✓	
$\mathcal{O}_{Q,q}^{3,1}$	✓	✓		$\mathcal{O}_{\phi D}$	✓	✓	✓
$\mathcal{O}_{tu}^{8}$	✓	✓		$\mathcal{O}_{\phi B}$		✓	
$\mathcal{O}_{tu}^{1}$	✓	✓		$\mathcal{O}_{\phi W}$		✓	
$\mathcal{O}_{td}^{8}$	✓	✓		$\mathcal{O}_{\phi WB}$	✓	✓	✓
$\mathcal{O}_{td}^{1}$	✓	✓		$\mathcal{O}_{e\phi,22}$		✓	
$\mathcal{O}_{Qu}^{8}$	✓	✓		$\mathcal{O}_{e\phi,33}$		✓	
$\mathcal{O}_{Qu}^{1}$	✓	✓		$\mathcal{O}_{u\phi,33}$	✓	✓	
$\mathcal{O}_{Qd}^{8}$	✓	✓		$\mathcal{O}_{d\phi,33}$		✓	
$\mathcal{O}_{Qd}^{1}$	✓	✓		$\mathcal{O}_{\phi e}$	✓	✓	✓
$\mathcal{O}_{tq}^{8}$	✓	✓	✓	$\mathcal{O}_{\phi b}$	✓	✓	✓
$\mathcal{O}_{tq}^{1}$	✓	✓	✓	$\mathcal{O}_{\phi d}$	✓	✓	✓
$\mathcal{O}_{\phi Q}^{1}$	✓	✓	✓	$\mathcal{O}_{\phi u}$	✓	✓	✓
$\mathcal{O}_{\phi Q}^{3}$	✓	✓	✓	$\mathcal{O}_{\phi q}^{(1)}$	✓	✓	✓
$\mathcal{O}_{\phi t}$	✓		✓	$\mathcal{O}_{\phi q}^{(3)}$	✓	✓	✓
$\mathcal{O}_{\phi tb}$	✓		✓	$\mathcal{O}_{\phi l}^{(1)}$	✓	✓	✓
$\mathcal{O}_{tB}$	✓		✓	$\mathcal{O}_{\phi l}^{(3)}$	✓	✓	✓
$\mathcal{O}_{tW}$	✓		✓	$\mathcal{O}_{ll}$	✓	✓	✓
$\mathcal{O}_{bW}$	✓		✓				
$\mathcal{O}_{tG}$	✓	✓					

Table 2.2: Process classes for the dimension six operators entering in the global SMEFT analysis. Left: List of operators for the top sector. Right: List of operators for the Higgs sector.

2.2 Electric dipole moments

With the SMEFT approach, we described an extension of the SM that helps address open BSM physics questions in a model-agnostic way. Another approach to address the baryon asymmetry as a first-order phase transition are EDMs. Similar to SMEFT, EDMs can be treated as low-energy EFT, based on Lagrangians of the weak- and hadronic-scale. For the baryon asymmetry, the Sakharov equations [16] can tell us precisely which structures are required: (i) C- and CP-violation, (ii) baryon number violation, and (iii) a deviation from thermal equilibrium. The first condition of the C- and CP-violation is especially interesting to us because it can be read off fundamental Lagrangians. The C-symmetry describes a symmetry under charge conjugation by switching all particles with their respective antiparticles. This holds for all charges relevant to the forces in the SM. As an example of processes in the SM violating the C-symmetry, one can name weak interactions, such as the β-decay [81]. Next, CP-symmetry is an extension of the C-symmetry, taking charge conjugation (C) and parity (P) into account. Parity describes the invariance under flipping the sign of all spatial coordinates. By combining C and P conservation, the CP-symmetry relates charge and spatial properties of the SM interactions and particles. Weak interactions, such as Kaon decay processes [82], violate this CP-symmetry in the SM. The origin of the CP-violation in the Kaon-decays is related to the effect of fermion

mixing among three generations in the SM. These fermion mixings are described by the Cabibbo-Kobayashi-Maskawa-matrix (CKM-Matrix), which not only contains the three mixing angles between the fermions, θ_{12}, θ_{13} and θ_{23}, but also a CP-violating complex phase δ [83]. Another source of CP-violation in the SM can arise from the QCD θ-term,

$$\frac{\theta}{32\pi^2} G^A_{\mu\nu} \tilde{G}^{A\mu\nu} \,, \tag{2.20}$$

with $G^A_{\mu\nu}$ the gluon field strength tensor and $\tilde{G}^{A\mu\nu}$ the dual field strength tensor. This term can contribute to CP-violation in strong interactions if it is non-zero. However, measurements of the neutron EDM being consistent with zero indicate a small value for θ, leading to the assumption that the CP violation in the QCD is too small to explain the baryon asymmetry reasonably [17–22].

EDM measurements provide a reliable way to detect CP violation since they provide a direct way to detect asymmetries between matter and antimatter. EDMs are a measure of separation between positive and negative electrical charges in a particle or system. If this electric dipole moment is non-zero, it indicates an imbalance in the charge distribution and, thus, a difference between the particle and antiparticle content of a system. In single particle systems, like the electron EDM or muon EDM, it is more complex to detect a non-zero EDM, as the electron itself is negatively charged and, thus, should not have any asymmetrical charge distribution. However, if there is a violation of the CP-symmetry, there might be an imbalance in the charge distribution, leading to a non-zero EDM measurement. The same holds for neutron EDM, which is the sum of the individual quarks and their charge. The measured EDMs from single particle and nucleon systems are small, additional systems, such as atoms and molecules, are used for EDM measurements. These are treated as asymmetrical systems consisting of smaller, individual parts, like electrons, protons, and neutrons. The CP-violating effect responsible for creating EDMs in these smaller parts can add up to a larger overall contribution to the EDM of the respective atomic or molecular system. With this, it is easier to detect the sum of multiple smaller parts being non-zero in more complex structures than directly measuring the EDM of the electron or other fundamental structures.

2.2.1 EDMs from Lagrangians

As seen in Fig. 2.1, EDM measurements can be directly connected to fundamental theories through WCs. Therefore, we start from a fundamental theory or Lagrangian, which is CP violating (CPV). After writing down these Lagrangians, they can be reduced to a set of low-energy effective Lagrangians. This procedure is equivalent to the top-down approach described in section 2.1.2. In the following global analysis, we start with a weak-scale and a hadronic-scale Lagrangian. The introduced WCs then can be matched onto the corresponding parameters in the low-energy theory, for example, C_S or C_T, which describe semileptonic interactions. These low-energy parameters and WCs can be determined from various measurements on an atomic and molecular level. Since fundamental Lagrangians describe these parameters, they can be determined using various measurements. Performing a global analysis allows us to include many different parameters and measurements simultaneously to encounter possible correlations and connections between these different parameters.

Next, we will discuss the building blocks of our global analysis, namely the underlying Lagrangians and the extraction of the different WCs used. Therefore, we consider the

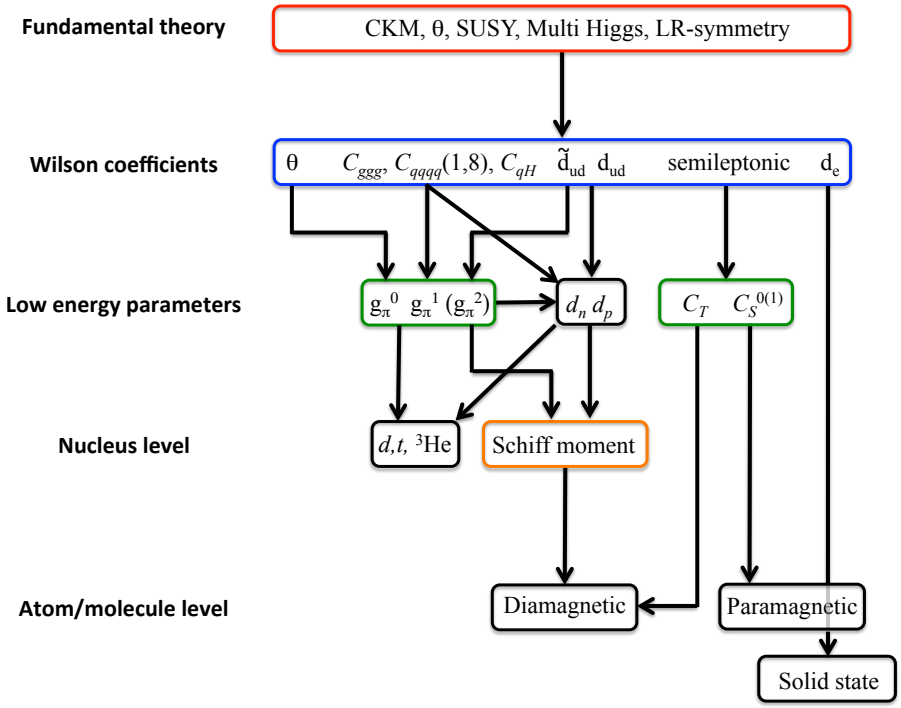

Figure 2.1: The connection from fundamental theories at high energy scales and their parameters to EDM measurements at low-energy scales. The displayed low-energy parameters are all constrained in the performed global analysis with the additional parameter C_P, defined in Eq. (2.38). The figure is taken from Ref. [26].

CP-violating weak-scale Lagrangian and the hadronic-scale Lagrangian. Later, these two Lagrangians are matched to create an overall parameter set.

This work focuses on combining different measurements and parameters in a global analysis and includes the effect of uncertainties at the hadronic scale. Thus, we will not provide a detailed analysis of CP-violating new physics at the electroweak scale and the renormalization group evolution from the GeV- to the electroweak scale.

Weak-scale Lagrangian

We start by discussing the modifications of the weak-scale Lagrangian and add a CP-violating behavior. For this weak-scale EFT we consider the SM leptons and quarks except the top quark, photons, and gluons as particles entering the Lagrangian. The CP-viloating behavior is introduced by the corresponding operators which appear in the Lagrangian as

$$\mathscr{L}_{\text{CPV}} = \mathscr{L}_{\text{CKM}} + \mathscr{L}_{\bar{\theta}} + \mathscr{L}_{\text{dipole}} + \mathscr{L}_{\text{Weinberg}} + \mathscr{L}_{\text{EFT}}\,, \tag{2.21}$$

where the CP violation coming from the neutrino sector is neglected [24]. The first term, $\mathscr{L}_{\text{CKM}}$, represents the CP violation from the δ-phase in the CKM matrix. The second

term, $\mathscr{L}_{\bar{\theta}}$, arises from a modified gluon field strength description, including the QCD θ-term

$$\mathscr{L}_{\bar{\theta}} = \frac{g_3^2}{32\pi^2}\bar{\theta}\,\mathrm{Tr}(G^{\mu\nu}\tilde{G}_{\mu\nu})\,, \tag{2.22}$$

where g_3 represents the strong coupling constant, $G^{\mu\nu}$ is the gluon field strength tensor, and $\widetilde{G}^{\mu\nu} = \epsilon^{\mu\nu\lambda\sigma}G_{\lambda\sigma}/2$ is the dual of the gluon field strength tensor. $\bar{\theta}$ describes the rescaled, CP-violating QCD parameter, where the bar notation indicates the included corrections from the quark mass matrix. In principle, $\bar{\theta}$ can be treated as a model parameter; however, by the agreement of the neutron EDM with zero, the value of $\bar{\theta}$ would be negligibly small. So, we can neglect this term in the global analysis.

The first two terms in the Lagrangian in Eq. (2.21) have mass dimension four, while the remaining terms are all higher dimensional and thus not part of a renormalizable extension. Continuing with the $\mathscr{L}_{\text{dipole}}$ term, it describes with d_f^E the electric dipole moments of fermions and d_q^C represents the chromoelectric dipole moments of quarks. Both appear at mass dimension five in the Larangian

$$\mathscr{L}_{\text{dipole}} = -\frac{i}{2}F^{\mu\nu}\sum_{f=q,\ell} d_f^E\,\left(\bar{f}\sigma_{\mu\nu}\gamma_5 f\right) - \frac{i}{2}g_3 G^a_{\mu\nu}\sum_{f=q} d_q^C\,\left(\bar{q}\sigma^{\mu\nu}\gamma_5 T^a q\right)\,. \tag{2.23}$$

The indices q and ℓ denote quarks and leptons of all three generations, and $F^{\mu\nu}$ is the electromagnetic field strength tensor. As a convention for the metric we choose $(1,-\mathbb{1})$ with $\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3$. T^a are the $SU(3)$ generators and $\sigma_{\mu\nu} = i\left[\gamma_\mu,\gamma_\nu\right]/2$ defines the fermion spins. The two parts of the Lagrangian $\mathscr{L}_{\text{dipole}}$ and $\mathscr{L}_{\text{Weinberg}}$, have mass dimension five, and the Weinberg part is defined as

$$\mathscr{L}_{\text{Weinberg}} = \frac{1}{3}d^G\,f_{abc}G^a_{\mu\nu}\widetilde{G}^{b\nu\rho}G^{c\,\mu}_{\rho}\,. \tag{2.24}$$

This part is again built on the gluon field strength tensor $G^a_{\mu\nu}$ and introduces the gluonic chromo-electric dipole moment d^G. Additionally, CP violation can also occur in terms of mass dimension six and higher, encoded in the term

$$\mathscr{L}_{\text{EFT}} = \sum_i \frac{C_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \mathcal{O}(\Lambda^{-3})\,, \tag{2.25}$$

which is only generated at higher energy scales $\Lambda \gg v$, with v the electroweak scale. The motivation behind this EFT expansion is the same as we used for the SMEFT framework [23, 24, 27–32, 84–87]. Relevant dimension-six operators generating EDMs include the following semileptonic and quark 4-fermion operators,

$$\begin{aligned}\mathscr{L}_{\text{EFT}} \supset & C_{\ell eqd}\,\left(\bar{L}^j e_R\right)\,\left(\bar{d}_R Q_j\right) + C^{(1)}_{\ell equ}\,\left(\bar{L}^j e_R\right)\epsilon_{jk}\left(\bar{Q}^k u_R\right) \\ & + C^{(3)}_{\ell equ}\,\left(\bar{L}^j\sigma_{\mu\nu} e_R\right)\epsilon_{jk}\left(\bar{Q}^k\sigma_{\mu\nu} u_R\right) + C^{(1)}_{quqd}\,\left(\bar{Q}^j u_R\right)\epsilon_{jk}\left(\bar{Q}^k d_R\right) \\ & + C^{(8)}_{quqd}\,\left(\bar{Q}^j T^a u_R\right)\epsilon_{jk}\left(\bar{Q}^k T^a d_R\right) + \text{h.c.}\,.\end{aligned} \tag{2.26}$$

Hadronic-scale Lagrangian

The challenge to describe EDMs by the Lagrangian, which is written down in Eq. (2.21), is that EDMs are measured far below the electroweak scale, where the propagating

degrees of freedom are leptons, non-relativistic nucleons $N = (p, n)^T$, with an average mass m_N, and pions $\vec{\pi} = (\pi^+, \pi^0, \pi^-)^T$ [23,26,27,33,84,86]. As convention for the isospin we use $\tau_3 |n\rangle = -|n\rangle$.

When we evolve the weak-scale Lagrangian to the experimentally more relevant GeV scale, only the lepton EDMs $d_\ell^E \equiv d_\ell$ in Eq. (2.23) remain unchanged. All other parameters are scale dependent, as the relation between charged leptons scales with the lepton mass. The muon and tauon EDMs can be included by factorizing them out from the hadronic-scale Lagrangian, but since the indirect constraints on EDMs are weaker than the direct ones and there is little interplay between model parameters, we can neglect them.

This leads to the following Lagrangian describing the EDMs at the relevant GeV-scale

$$\mathscr{L}_{\text{had}} \supset \mathscr{L}_N + \mathscr{L}_{\pi N} + \mathscr{L}_{eN} - \frac{i}{2} F^{\mu\nu} d_e \, \bar{e} \sigma_{\mu\nu} \gamma_5 e \,, \tag{2.27}$$

where we include the observable for the electron EDM d_e directly. $\mathscr{L}_N$, as the first term in the Lagrangian, is defined as

$$\mathscr{L}_N = -2\bar{N} \left[d_p \frac{1+\tau_3}{2} + d_n \frac{1-\tau_3}{2} \right] S_\mu N v_\nu F^{\mu\nu} \,, \tag{2.28}$$

where S_μ and v_ν are the nucleon's spin and velocity, and the choice of d_N is defined by the isoscalar and isovector contributions. Next, the second term $\mathscr{L}_{\pi N}$ in the Lagrangian describes the interactions between pions and nucleons,

$$\begin{aligned} \mathscr{L}_{\pi N} =& \bar{N} \Big[g_\pi^{(0)} \vec{\tau} \cdot \vec{\pi} + g_\pi^{(1)} \pi^0 + g_\pi^{(2)} \left(3\tau_3 \pi^0 - \vec{\tau} \cdot \vec{\pi} \right) \Big] N \\ &+ C_1 \, \left(\bar{N} N \right) \, \partial_\mu \left(\bar{N} S^\mu \bar{N} \right) + C_2 \, \left(\bar{N} \vec{\tau} N \right) \cdot \partial_\mu \left(\bar{N} S^\mu \bar{N} \vec{\tau} \right) + \cdots \,, \end{aligned} \tag{2.29}$$

with τ being the Pauli matrices and the ellipses indicating higher-order interactions, such as terms including more than one pion. These higher-order terms are neglected because of their suppression compared to the leading-order terms. Also, the contribution involving $g_\pi^{(2)}$, representing the isotensor contribution, is suppressed relative to $g_\pi^{(0,1)}$, the corresponding isoscalar and isovector contributions, by one order in the chiral expansion [88, 89]. The second line in Eq. (2.29) can be neglected because the shown chiral expansion in terms of C_i is suppressed and only enters at NNLO. Finally, the second term in Eq. (2.27) describes the effective interactions between electrons and nucleons, corresponding to the higher-dimensional operators in Eq. (2.26). These contributions can be organized according to their tensor structure, isospin character, and their dependence on the electron and nucleon fields, and spins [26, 27]:

$$\begin{aligned} \mathscr{L}_{eN} =& -\frac{G_F}{\sqrt{2}} \, (\bar{e} i \gamma_5 e) \, \bar{N} \left(C_S^{(0)} + C_S^{(1)} \tau_3 \right) N \\ &+ \frac{8 G_F}{\sqrt{2}} \, v_\nu \, (\bar{e} \sigma^{\mu\nu} e) \, \bar{N} \left(C_T^{(0)} + C_T^{(1)} \tau_3 \right) S_\mu N \\ &- \frac{G_F}{\sqrt{2}} \, (\bar{e} e) \, \frac{\partial^\mu}{m_N} \left[\bar{N} \left(C_P^{(0)} + C_P^{(1)} \tau_3 \right) S_\mu N \right] \,, \end{aligned} \tag{2.30}$$

where S, P, T denote the scalar, pseudoscalar and tensor contributions, 0 denotes the isoscalar part of that contribution and 1 the isovector part.

Taking all observables from the hadronic scale Lagrangian from Eq. (2.27) into account,

we end up with eleven independent Lagrangian parameters for the global analysis, namely

$$\left\{ d_e, C_S^{(0,1)}, C_T^{(0,1)}, C_P^{(0,1)}, g_\pi^{(0)}, g_\pi^{(1)}, d_{n,p} \right\} . \tag{2.31}$$

2.2.2 Matched Lagrangians

The set of model parameters listed in Eq. (2.31) can be further simplified by matching the semileptonic part of the hadronic-scale Lagrangian at the GeV-scale to the 4-fermion interaction in Eq. (2.26) of the weak-scale Lagrangian. Therefore, we evaluate both Lagrangians for external nucleons. By doing so, the light quarks inside these nucleons are related to $\mathscr{L}_N$ via

$$\begin{aligned}
g_S^{(0)}\ \bar\psi_N\psi_N &= \frac{1}{2}\langle N \left|\bar u u + \bar d d\right| N\rangle \\
g_S^{(1)}\ \bar\psi_N\tau_3\psi_N &= \frac{1}{2}\langle N \left|\bar u u - \bar d d\right| N\rangle \\
g_T^{(0)}\ \bar\psi_N\sigma_{\mu\nu}\psi_N &= \frac{1}{2}\langle N \left|\bar u \sigma_{\mu\nu} u + \bar d \sigma_{\mu\nu} d\right| N\rangle \\
g_T^{(1)}\ \bar\psi_N\sigma_{\mu\nu}\tau_3\psi_N &= \frac{1}{2}\langle N \left|\bar u \sigma_{\mu\nu} u - \bar d \sigma_{\mu\nu} d\right| N\rangle \\
g_P^{(0)}\ \bar\psi_N\gamma_5\psi_N &= \frac{1}{2}\langle N \left|\bar u \gamma_5 u + \bar d \gamma_5 d\right| N\rangle \\
g_P^{(1)}\ \bar\psi_N\gamma_5\tau_3\psi_N &= \frac{1}{2}\langle N \left|\bar u \gamma_5 u - \bar d \gamma_5 d\right| N\rangle .
\end{aligned} \tag{2.32}$$

These relations now define the scalar, pseudoscalar, and tensor contributions of the nucleon form factors $g_{S,T,P}^{(0,1)}$. Now, the hadronic-scale WCs written down in Eq. (2.30) can be matched to the SMEFT WCs from Eq. (2.26) as [24, 26],

$$\begin{aligned}
C_S^{(0)} &= -g_S^{(0)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell edq} - C_{\ell equ}^{(1)}\right) & C_S^{(1)} &= \ \ g_S^{(1)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell edq} + C_{\ell equ}^{(1)}\right) \\
C_T^{(0)} &= -g_T^{(0)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell equ}^{(3)}\right) & C_T^{(1)} &= -g_T^{(1)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell equ}^{(3)}\right) \\
C_P^{(0)} &= \ \ g_P^{(0)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell edq} + C_{\ell equ}^{(1)}\right) & C_P^{(1)} &= -g_P^{(1)}\frac{v^2}{\Lambda^2}\,\mathrm{Im}\left(C_{\ell edq} - C_{\ell equ}^{(1)}\right) .
\end{aligned} \tag{2.33}$$

Here, the six couplings $C_{S,T,P}^{(0,1)}$ describing the scalar, tensor, and pseudoscalar semileptonic interactions with their respective isoscalar and isovector contribution, are expressed in terms of only three SMEFT Wilson coefficients at dimension six, implying

$$\frac{C_P^{(0)}}{g_P^{(0)}} = \frac{C_S^{(1)}}{g_S^{(1)}} \qquad \frac{C_T^{(0)}}{g_T^{(0)}} = \frac{C_T^{(1)}}{g_T^{(1)}} \qquad \frac{C_S^{(0)}}{g_S^{(0)}} = \frac{C_P^{(1)}}{g_P^{(1)}} . \tag{2.34}$$

These relations link the $C_{S,P,T}^{(0)}$ to the $C_{S,P,T}^{(1)}$ coefficients using $g_{S,T,P}^{(0,1)}$. With this, we can reduce the number of semileptonic parameters down to three independent ones that actually enter the hadronic-scale global analysis. We choose them as $C_{S,T,P}^{(0)}$ and combine them using the known ratios of hadronic matrix elements to construct the full Lagrangian. However, we still preserve the full dependence on the $C_{S,P,T}^{(1)}$ coefficients, even though there might be isospin-violating effects. Furthermore, this relation also holds when, for the included experimental systems, the $C_S^{(1)}$ might differ significantly. This effect will be discussed later.

Additionally to the light quark contributions described by Eq. (2.33), we have to adopt the relations in Eq. (2.34) to also consider the heavy quark contributions. These contributions are hidden in the nucleon form factors, which are renormalized at an appropriate mass scale.

Starting from the relations for $C_S^{(0,1)}$, the effective parameter combining the isoscalar and isovector parts of the scalar contributions is independent of the nuclear spin. This makes the implementation fairly straightforward,

$$\begin{aligned} C_S &= C_S^{(0)} + \frac{Z-N}{Z+N}\, C_S^{(1)} \\ &= C_S^{(0)} + \frac{Z-N}{Z+N} \frac{g_S^{(1)}}{g_P^{(0)}}\, C_P^{(0)} \qquad \text{with} \qquad \frac{g_S^{(1)}}{g_P^{(0)}} \approx 0.1\,. \end{aligned} \tag{2.35}$$

In the second line, we replace, based on Eq. (2.33), $C_S^{(1)}$ with $C_P^{(0)}$. Based on the small isospin violation of the nucleon matrix element, $g_S^{(1)}$ is suppressed relatively to $g_S^{(0)}$. If we apply that knowledge to Eq. (2.35), one can argue that $C_S^{(1)} \ll C_S^{(0)}$. Additionally, having a look at all heavy nuclei in the atomic and molecular systems for which EDMs have been measured, it is recognizable that the isoscalar and isovector contributions always occur in approximately the same ratio, $(Z-N)/(Z+N) \approx -0.2$. Based on these findings, it is arguable that the parameter $C_S^{(0)}$ might be system-independent, but we do not rely on this assumption and still include $C_S^{(0)}$ as a free parameter in the global analysis.

Next, we move on by relating the pseudoscalar and tensor parts of the semileptonic interactions. There, we start from the linear combination of nucleons, namely $C_{P,T}^{(n,p)} = C_{P,T}^{(0)} \mp C_{P,T}^{(1)}$, with the upper sign reflecting n, accordingly to our chosen isospin convention. Based on that relation, the coefficients for any given nucleus are constructed by the sum over spins of the constituent nucleons, where $\langle \sigma_{p,n} \rangle$ is the expectation value of these spins, evaluated via Pauli operators for the measured nuclear state [90, 91]

$$C_{P,T} = \frac{C_{P,T}^{(n)} \langle \sigma_n \rangle + C_{P,T}^{(p)} \langle \sigma_p \rangle}{\langle \sigma_n \rangle + \langle \sigma_p \rangle}\,. \tag{2.36}$$

For C_T, we can see from Eq.(2.33) that the isoscalar and isovector couplings differ only through the corresponding nucleon form factors. These are calculated with small theoretical uncertainties in lattice QCD [92], which allows us to write

$$C_T = \left(1 - \frac{g_T^{(1)}}{g_T^{(0)}} \frac{\langle \sigma_n \rangle - \langle \sigma_p \rangle}{\langle \sigma_n \rangle + \langle \sigma_p \rangle}\right) C_T^{(0)} \qquad \text{with} \qquad \frac{g_T^{(1)}}{g_T^{(0)}} \approx 1.7\,. \tag{2.37}$$

Similarly, for C_P, it can be shown that

$$C_P = C_P^{(0)} - \frac{g_P^{(1)}}{g_S^{(0)}} \frac{\langle \sigma_n \rangle - \langle \sigma_p \rangle}{\langle \sigma_n \rangle + \langle \sigma_p \rangle}\, C_S^{(0)} \qquad \text{with} \qquad \frac{g_P^{(1)}}{g_S^{(0)}} \approx 20.2\,. \tag{2.38}$$

Coming from the derivations of $C_{S,P,T}$, we are left now with $g_P^{(0,1)}$. Starting by the derivation of $g_P^{(1)}$, we follow Ref. [24]. Thus, we only consider the first generation of

quarks as the relevant light ones, in a way that $g_P^{(1)}$ is dominated by the π-pole contribution

$$g_P^{(1)} = \frac{g_A \bar{m}_N}{\bar{m}} \frac{m_\pi^2}{m_\pi^2 - q^2} + \text{heavy quarks} \qquad \bar{m} = \frac{m_u + m_d}{2} . \tag{2.39}$$

With $\bar{m}_N \approx 940$ MeV being the average nucleon mass, $\bar{m}$ the average light quark mass, and g_A is the isovector part of the axial vector coupling [93]. Next, the coupling $g_P^{(0)}$ appearing in Eq. (2.35), involves the isoscalar axial coupling $g_A^{(0)}$, which is obtained from the sum of the light quark axial charges [94]. We extend this light quark contribution to $g_P^{(0)}$ by also allowing a light s-quark. Now the π-pole dominance has to be replaced by an octet η-pole with a modified average light quark mass m^*,

$$g_P^{(0)} = \frac{g_A^{(0)} \bar{m}_N}{m^*} \frac{m_\eta^2}{m_\eta^2 - q^2} + \text{heavy quarks} \qquad m^* = \frac{\bar{m} + 2m_s}{3} . \tag{2.40}$$

Note, that the heavy quarks enter differently in $g_P^{(1)}$ compared to $g_P^{(0)}$, due to the different poles we consider. This difference can be derived using the $U(1)_A$ axial anomaly together with the divergence of the anomaly-free axial current $J_{\mu 5}^q = \bar{q}\gamma_\mu\gamma_5 q$ for all quarks q [94–97]. However, these represent only a relatively minor contribution to $g_P^{(1)}$, but since the η-pole is suppressed by a factor $m^*/\bar{m}$ compared to the π-pole, its contribution to $g_p^{(0)}$ is approximately at the same level as the light quark contribution.

As a final simplification, not really necessary, but still reducing the parameter set by one and removing a poorly constrained direction in model space, we assume

$$d_p \approx -d_n . \tag{2.41}$$

To justify this choice, we have to look at the structures of the atoms and molecules of the different measurements. These measurements will be discussed in more detail in Chapter 5.1. Considering the nuclei of the measured closed-shell system, these are typically dominated either by a valence proton (see TlF) or a valence neutron (all others). Also, one can assume that the short-range nucleon EDMs are dominated by their isovector contribution. In the sense of global analyses, this assumption can be relaxed since we also include theory uncertainties concerning the contributions of all nucleons to the overall EDM. In addition, with the present experimental limits, only TlF has a leading sensitivity for d_p.

With these simplifications, as described in this section, the set of low-energy parameters given in Eq.(2.31) reduces to

$$c_j \in \left\{ d_e, C_S^{(0)}, C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n \right\} . \tag{2.42}$$

These seven Lagrangian parameters define the model parameter for our global EDM analysis.

Finally, the different measurements we will consider in the global analysis will be discussed in Chapter 5.1. With this, we define the bases for global analyses in the SMEFT and EDM sectors. These results provide the theoretical framework needded to perform the global analyses using SFitter as tool. This will also include the discussion of different uncertainty treatments on model parameters, such as in the EDM analysis, and on experimental data, which plays a role in both analyses.

Chapter 3

SFitter as a Tool for Global Analyses

In the previous chapter, we introduced the concept of EFTs in the high-energy and low-energy limits, using SMEFT and the weak-scale EFT for describing EDMs. These EFTs are powerful frameworks for discussing new physics effects in an (almost) model independent and global approach. The next challenge is now to analyze the data and evaluate the framework provided by the model to determine the values of the corresponding WCs. We choose to use the SFitter framework [98] as analysis tool of our choice to answer these questions. SFitter was developed as a multi-purpose analysis tool for various topics, for example supersymmetric (SUSY) dark matter [99–104] and Higgs searches [105–107], but it also serves as tool for global SMEFT [1, 75, 79, 108–111] and EDM [2] analyses. Its original prupose is analysing different searches in particle physics, especially to test different SUSY models. However, it has recently been adopted to include other physical models containing multiple measurements and parameters. This is possible due to its flexible and adaptive framework, making including new processes and models fairly easy. SFitter is based on the principle of a basic likelihood analysis, where the predictions and corresponding measurements are compared to determine the overall likelihood. This likelihood is then computed for one point in the model or parameter space at a time and later combined into a global approach. The used experimental data in this analysis is implemented by including correlations and different kinds of uncertainties that occur when performing experiments and matching them to theoretical predictions. Including these different measurements, uncertainties, and correlations gives us a combination of individual likelihoods when considering multiple parameters, which later get combined into a full, exclusive likelihood.

When we perform a global analysis in the SMEFT sector, the used model parameters are the WCs obtained from higher-order terms in the SMEFT Lagrangian written in Eq. (2.6). The data is obtained from different collider measurements and experiments. This includes data mostly from the ATLAS and CMS experiments, but also EWPO measurements from the Large Electron–Positron Collider (LEP). Overall, we include total and differential cross-sections, rate and signal-strength measurements.
The experimental data in the EDM sector consists of different EDM measurements, including neutron EDM, molecules like ThO, HfF, and TlF, and atomic EDM measurements, such as Hg, Ra, or Tl. The model parameters listed in Eq. (2.42) are obtained from the matching of the weak-scale Lagrangian from Eq. (2.21) to the hadronic-scale Lagrangian in Eq. (2.27).

Technically, the global analysis is performed by sampling around an initial point of interest followed by mapping out the whole phase space using an MCMC algorithm. The values of the considered WCs are a priori unknown, and need to be determined from

experimental data. Therefore, we consider all the relevant operators and WCs for the included measurements. With this procedure, we encounter a problem: We have a model providing predictions for physical parameters of a set of free parameters and we have a set of experimental data. Thus, we must find a way to combine both aspects to get reasonable, well-defined results that best describe the data. However, uncertainties and correlations still have to be included in the parameter estimation. Therefore, the overall fitting can be described as a two-step process:

1. Define a measure for the goodness of the fit.
2. Maximize or minimize this measure.

This chapter is structured as follows: First, we will focus on constructing the likelihood, including different kinds of uncertainties and the generalization towards a combined likelihood. Next, we will discuss the inclusion of correlations in the likelihood, how to evaluate this likelihood, and which statistical test is used. Lastly, we will discuss the extraction of the limits on the actual parameters for a profiling and marginalization approach.

3.1 Constructing the likelihood

To find the best model parameter configuration for the given experimental data, first, one has to define a measure for the quality of the fit. We will see that the likelihood approach serves as the most powerful and versatile ansatz to determine the goodness of the fit and demonstrate how to construct the likelihood and generalize it to multiple parameters. Next, in Sec. 3.2, we will discuss the evaluation and optimization of this likelihood to examine the best-fit point of the model.

When the same experiment is performed several times, we would expect different outcomes every time, either due to the measuring devices' imprecision or the system's stochastic nature. A probability density function (PDF) p of the measured parameter x describes the expected distribution of those different outcomes. Naturally, the PDF p needs to fulfill the condition

$$\int p(x)\mathrm{d}x = 1 \,, \tag{3.1}$$

and being larger than zero for all possible and allowed values for x. The above equation misses a crucial part, the expected distribution is not only dependent on the observable x but also on the model parameter α. Thus, the PDF should be written as $p(x|\alpha)$ in general. For clarity, we will assign Latin letters to the measurements and Greek ones to the model parameters. The model is defined over all predictions encoded by α and the assumptions we make on the PDF. The PDF also depends on the experimental results and the uncertainties affecting the experimental outcome. Therefore, if we want to consider the full model, we have to take an effective description of the experimental values depending on the observables x into account and their influence on the assumed PDF shape. We distinguish between three PDF distributions and thus three different uncertainty types:

1. **Poissonian distribution (statistical uncertainties)**
   Uncertainties following a Poissonian distribution are intrinsic for each measurement. These might vanish for an infinite number of measurements performed or events

recorded and thus are also called statistical uncertainties. It is defined as the probability of observing d events, when expecting ν_d events. The PDF for the Poisson distribution is given by

$$p_{\text{Pois}}(d|\nu_d) = \frac{(\nu_d)^d}{d!} e^{-\nu_d},\ \nu_d > 0\ , \tag{3.2}$$

with the variance given by $\sigma_d^2 = \nu_d$.

2. **Gaussian distribution (systematic uncertainties)**
In the limit of large d and ν_d, the Poisson distribution in Eq. (3.2) converges towards a Gaussian distribution

$$p_{\text{Gauss}}(d|\nu_d) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(d-\nu_d)^2}{2\sigma_d^2}\right)\ , \tag{3.3}$$

which is used to describe systematic uncertainties. Additionally, to statistical uncertainties, systematic ones arise from external factors and will not vanish within the limit of repeating an experiment an infinite number of times. The uncertainty on the luminosity, photon reconstruction, tagging, or the energy scales are considered as systematic uncertainties. Since we usually have large amounts of data to extract, we expect them to follow a Gaussian shape, but this assumption does not necessarily have to be true. In principle, assuming a Gaussian distribution here is only an approximation.

3. **Flat distribution (Theory uncertainties)**
Lastly, we have to consider uncertainties arising from theoretical predictions. Unlike experimental uncertainties, these do not follow a specific distribution. Rather, the outcome of a measurement is equally likely within the range σ_d around the central value ν_d and we do not expect any tails in the distribution, leading to $d \in [\nu_d - \sigma_d, \nu_d + \sigma_d]$. Two examples for a theoretical uncertainty are either scale uncertainties or the uncertainties on the semileptonic and hadronic parameters in the EDM framework. Regarding scale uncertainty, a central scale μ is usually varied by a factor 2. This leads to a possible spread between $\mu/2$ and 2μ which has no statistical meaning. All outcomes in this area with a width of $2\sigma_d$ are equally probable and there is no reason to favor one outcome over the other. But beyond those boundaries, the assumption is not valid, leading to a distribution without any tails. Similar to calculating the hadronic and semileptonic parameters of the EDM Lagrangian, they depend on various choices regarding masses and spin states, with this all outcomes are equally likely. For these reasons, we assume a flat distribution with a box-shape, which is realised using a Heaviside function Θ

$$p_{\text{Flat}}(\nu_d|\nu_d^{\text{min}}, \nu_d^{\text{max}}) = \frac{1}{2\sigma_d}\Theta(\nu_d - \nu_d^{\text{min}})\Theta(\nu_d^{\text{max}} - \nu_d)\ , \quad \sigma_d = (\nu_d^{\text{max}} - \nu_d^{\text{min}})\ . \tag{3.4}$$

For a fixed measurement x the PDF $p(x|\alpha)$ is called a likelihood $L(\alpha) = L(x|\alpha) = p(x|\alpha)$. The Likelihood $L(\alpha)$ is not a probability in terms of α, since it is not normalized. For convenient reasons it is common to work with the negative log-likelihood $-2\log L(\alpha)$ instead of $L(\alpha)$. The product of several likelihoods translates to a sum for the log-likelihood approach, and all normalization factors become additive constants.

3.1.1 Combining likelihoods

Normally, we have more than one measurement, model parameter, and uncertainty. How do we combine these different likelihoods into one likelihood we can implement into SFitter? For n independent and identically distributed measurements x in a data set D the combined likelihood reads

$$L(\alpha) = L(D|\alpha) = \prod_{i=1}^{n} p(x_i|\alpha) \; . \tag{3.5}$$

This decomposition can be used to combine likelihoods from different measurements. Resulting in an exclusive likelihood, which is a function dependent on the model parameters α_x, nuisance parameters for systematic θ_{syst} and theory θ_{theo} uncertainties, and the expected background b of the measurement. Nuisance parameters are additional parameters, which are relevant for an accurate description of the model without being of interest in our fit. With this we can start from the exclusive likelihood for one channel, and then generalize it to multiple channels.

The exclusive likelihood for one channel is constructed from different building blocks multiplied with each other

$$\begin{aligned} L_{\text{excl}}(\alpha_x, \theta_{\text{syst}}, \theta_{\text{theo}}, b) &= \text{Pois}(d|m(\alpha_x, \theta_{\text{syst}}, \theta_{\text{theo}}, b)) \\ &\quad \times \text{Pois}(b_{\text{CR}}|b\ k) \prod_i \mathcal{N}_{\theta_{\text{syst},i},\sigma_i}(0) \prod_j \mathcal{F}_{\theta_{\text{theo},j},\sigma_j}(0) \; . \end{aligned} \tag{3.6}$$

The first factor in the exclusive likelihood represents the expected contribution from the considered model m depending on the different parameters the model contains, like α, the background b and the respective uncertainties for a given observed data d. The second term shows the contribution from the expected background b, given the observed background in the control region (CR) b_{CR} and an interpolation factor k between the CR and signal region (SR). The signal region is defined as the part of the data set where the signal is expected to appear. In contrast the control region is dominated by the background and used to estimate and validate the background in the SR. Due to their statistical nature, both contributions are best described by a Poissonian distribution. The last two factors given in Eq. (3.6), represent the contributions of nuisance parameters for systematic θ_{syst} and theory θ_{theo} uncertainties. They should be centered around zero with a given half-width σ. As mentioned before, in SFitter the systematic uncertainties should follow a Gaussian and the theory uncertainties should follow a flat distribution.

As derived in Eq. (3.6) this likelihood for one channel can be generalized to N channels by replacing

$$\begin{aligned} \text{Pois}(d|m)\text{Pois}(b_{\text{CR}}|b\ k) &\to \prod_l \text{Pois}(d_l|m_l)\text{Pois}(b_{\text{CR}_l}|b_l\ k_l) \\ \mathcal{N}_{\theta_{\text{syst},i},\sigma_i}(0) &\to \mathcal{N}_{\vec{\theta}_{\text{syst},i},\Sigma_i}(\vec{0}) \\ \mathcal{F}_{\theta_{\text{theo},j},\sigma_j}(0) &\to \prod_l \mathcal{F}_{\theta_{\text{theo},lj},\sigma_{lj}}(0) \; . \end{aligned} \tag{3.7}$$

These generalizations imply uncorrelated theory uncertainties and correlated systematic uncertainties. The correlations between systematic uncertainties are encoded using an N-dimensional Gaussian with Σ_i as covariance matrix with uncertainties in a category i entering different channels l. In our global analyses, we either fully correlate or

uncorrelate systematic uncertainties. This choice is justified by the fact, that the appearing correlations are negligible or fully correlated, as we show in Ch. 4.

In SFitter there are two possible ways to evaluate the constructed likelihood: By profiling over the likelihood or by applying Bayesian marginalization. Section 3.2.2 will present more details on these two methods. In the context of marginalization, the second line in Eq. (3.6) can be understood as priors, which encode our prior knowledge or belief on the expected background b and nuisance parameters θ_{syst} and θ_{theo}. While for the profiling approach, these assumptions should be understood as PDFs constructed from auxiliary measurements.

Starting from the constructed exclusive likelihood in Eq. (3.6), and after generalizing it to multiple channels in Eq. (3.7) we can move one to construct either the profiled or marginalized likelihood. The profiled likelihood is obtained by taking the maximum of the likelihood varying the background b and the nuisance parameters θ_{syst} and θ_{theo}

$$L_{\text{profile}}(\alpha_x) = \max_{b,\theta_{\text{syst}},\theta_{\text{theo}}} L_{\text{excl}}(\alpha_x, \theta_{\text{syst}}, \theta_{\text{theo}}, b) \; . \tag{3.8}$$

On the other hand, the marginalized likelihood is obtained by integrating over these parameters

$$L_{\text{marg}}(\alpha_x) = \int \prod_i \mathrm{d}\theta_{\text{syst},i} \int \prod_j \mathrm{d}\theta_{\text{theo},j} \int \mathrm{d}b L_{\text{excl}}(\alpha_x, \theta_{\text{syst}}, \theta_{\text{theo}}, b) \; . \tag{3.9}$$

From a theoretical point of view, the only difference between both methods is shown in Eqs. (3.8) and (3.9) by either computing the maximum of the nuisance parameter or integrating over those, however, there are more assumptions to take care of in the practical implementation. To make the marginalization approach in highly dimensional spaces efficient, we must find a way to approximate the integrals. Although many of the integrals written down in Eq. (3.9) can be solved analytically, generalizing it to N channels still requires us to solve a N-dimensional integral. Therefore, we choose to solve them numerically. These and other assumptions for the marginalization treatment are discussed in detail in Ref. [75]. Regarding the profiling treatment, more information on handling uncertainties and different channels can be found in Ref. [112, 113].

3.1.2 Introducing correlations

As mentioned for the generalization of the exclusive likelihood in Eq. (3.7), we can consider correlations between systematic uncertainties.

For LHC measurements, systematic uncertainties are often estimated for single measurements based on the same secondary measurements. The uncertainty on the luminosity, for example, is determined once and then used for all LHC experiments. Some detector effects from CMS or ATLAS are also estimated and calibrated once and then applied to every measurement. We consider the correlations of two experimental channels, i and j, as a dimensionless measure. This measure describes the dependence of both channels on each other. It is given by a quadratic relationship through the correlation matrix C

$$C_{ij} = \text{corr}(x_i, x_j) = \frac{V_{ij}}{\sigma_i \sigma_j} = \frac{E[(x_i - \bar{x_i})(x_j - \bar{x_j})]}{\sigma_i \sigma_j} \; , \tag{3.10}$$

where V denotes the covariance, calculated from the expectation value E of the deviations from the measurements x from their mean $\bar{x}$. From Eq. (3.10) we can directly see, that the correlation matrix has to be symmetric with 1 as its diagonal entries. The off-diagonal entries are the correlation coefficients $\rho_{ij} = C_{ij}$. In practice, these correlation coefficients ρ_{ij} are often unknown and need to be estimated from the individual uncertainty sources. As we only correlate systematic uncertainties, we can replace V_{ij} in Eq. (3.10) by $V_{ij}^{\text{syst}} = \sigma_{i,\text{syst}}\sigma_{j,\text{syst}}\rho_{ij,\text{syst}}$ and use the full experimental uncertainty consisting of systematic and statistical uncertainties for $\sigma_i\sigma_j$:

$$C_{ij} = \frac{\sum_{\text{syst}} \sigma_{i,\text{syst}}\sigma_{j,\text{syst}}\rho_{ij,\text{syst}}}{\sigma_{i,\text{exp}}\sigma_{j,\text{exp}}} , \qquad \text{with} \quad \sigma^2_{i,\text{exp}} = \sum_{\text{syst}} \sigma^2_{i,\text{syst}} + \sum_{\text{Pois}} \sigma^2_{i,\text{Pois}} . \tag{3.11}$$

In SFitter we assume all systematic uncertainties to be fully correlated between measurements of the same experiment, the only exception is the luminosity, which is correlated between the ATLAS and CMS experiment. To be more specific, we choose $\rho_{ij} = 0.99$, to ensure the invertability of the final correlation matrix is still given.

Finally, the full log-likelihood, or equivalently, the χ^2 is computed as

$$-2\log L = \chi^2 = \chi_i^T C_{ij}^{-1} \chi_j , \tag{3.12}$$

with $\chi_i^2 = -2\log L_i$ for each individual channel i and thus $\chi_i = \sqrt{\chi_i^2}$.

3.2 Likelihood mapping and evaluation

Next, we are interested in mapping the likelihoods in the regions around the best fit point, the point with the highest likelihood, and how to evaluate the results.

To perform this mapping of the phase space of the likelihood, we need two ingredients. First, a reliable method to compare two likelihood points, a statistical test, and second, an efficient sampling of the parameter space, where we use the MCMC approach.

3.2.1 Likelihood ratio as statistical test and MCMC

To compare two models or hypotheses with each other, we need a statistical test. This test serves as a measure for the goodness of the fit, where a test model m_{test} is compared to a reference or baseline model m_0 given the experimental data d. In SFitter the likelihood ratio is chosen as the statistical test method. This method was proven by Neyman and Pearson [114] to be the most efficient one. Therefore, we consider a zero hypothesis m_0 and a test hypothesis m_{test} and compute the likelihood ratio as

$$t(m) = \frac{L(m_{\text{test}}|d)}{L(m_0|d)} . \tag{3.13}$$

Where in the case of setting limits, the point for the zero hypothesis H_0 is chosen as the best fit point of the model $m_0 = m_{\text{best}}$, such that $L(m_{\text{best}}|d) = \max_m L(m|d)$. When it is implemented in a MCMC setting to determine the next point, H_0 is the current model point of the chain $m_0 = m_{\text{current}}$.

Using the likelihood ratio as test statistics also has another benefit. According to Wilk's theorem [115], the likelihood ratio follows a χ^2-distribution for a large sample size. The

degrees of freedom of the χ^2-distribution are precisely the number of parameters α in our model configuration m [115, 116]. This allows us to use the known χ^2 properties on setting limits to evaluate the parameters in our analysis.

Furthermore, we need an efficient way to map the phase space region around the best fit point accordingly. We use a MCMC algorithm to map the phase space region around this best fit point. With this, a test point m_{test} is obtained by sampling from either a Gaussian, Breit-Wigner or flat distribution, centered at the current point m_{current}. Thus, the acceptance of the test point only depends on the present point and not any previous points of the chain [117]. Then, the acceptance of m_{test} depends on the outcome of the likelihood ratio from Eq. (3.13), comparing the test point with the current point. If

$$\frac{L(m_{\text{test}}|d)}{L(m_{\text{current}}|d)} > 1 \ , \tag{3.14}$$

the test point is accepted and added to the chain as new current point $m_{\text{next}} = m_{\text{current}}$. If the ratio is smaller than one, the m_{test} only gets accepted in the case of

$$\frac{L(m_{\text{test}}|d)}{L(m_{\text{current}}|d)} \geq r \ , \tag{3.15}$$

with $r \in [0, 1]$, a random value sampled from a uniform distribution [118, 119]. Otherwise m_{test} gets rejected and the current point remains the next $m_{\text{next}} = m_{\text{current}}$.

This algorithm is called the Metropolis algorithm [120], which ensures that the probability of the test point to be selected is

$$\min\left(1, \frac{L(m_{\text{test}}|d)}{L(m_{\text{current}}|d)}\right) \ . \tag{3.16}$$

Starting from a random test point, the chain prefers new points with higher likelihoods than the previous point. This drives the MCMC to the region of the highest likelihood. However, by introducing r as a rate to accept points with a lower likelihood, the algorithm is also able to map out regions around the maxima. It can also be used to explore more complex structures. Suppose all the parameters of the MCMC are fine-tuned accordingly. In that case, it scales linearly with the dimensionality of the parameter space, making it computationally less expensive than other methods, such as using a grid. To suppress the influence of the initial conditons on the outcome of the chain we introduce a burn-in phase, which deletes the start of the chain up to a certain percentage of points. For the profiling approach this limit is usually set to 1%, while for marginalization it is around 10%. The higher percentage for the marginalization is justified by the integration over the likelihood, which requires an accurate representation of the phase space density.

3.2.2 Extracting parameter limits

Next, we discuss extracting parameter limits from the likelihood and the MCMC results. Starting from the map of the profiled and marginal likelihood in Eqs. (3.8) and (3.9), we want to extract the limits on the corresponding parameters of our model. Therefore, we profile or marginalize over all but the one or two parameters we want to constrain. From here onwards, the treatment is different for either profiling or marginalization, so we discuss them separately.

For the profile likelihood, the proper way to define these intervals is to use the Feldman-Cousins approach, which is based on Neyman's construction of the likelihood ratio as an ordering rule. However, this approach is cumbersome and numerically challenging. We use likelihood based intervals as an alternative. By caluclating the χ^2 value in Eq. (3.12), we can set the limits on the χ^2 distribution by calculating parameter values which exceed a certain threshold. This threshold depends on the number of DOFs and the confidence level (CL) one wants to set. The values for either one or two DOFs and the corresponding CLs are given in Tab. 3.1.

Extracting limits from the marginalization approach is slightly different. We start by identifying the global maximum of our likelihood and then integrate around that point. While integrating, we choose integral borders on iso-likelihood contours. If there are additional peaks, their likelihood is computed only if they exceed a certain threshold, and then only that part exceeding the threshold is integrated. We stop integrating to determine the interval's limits when the integral value is larger than 68% or 95% of the full integral.

The Bayesian counterpart to confidence intervals is called credible intervals and defined by the marginalization approach. Confidence and credible intervals are two different things, and thus, the limits do not have to match. However, to make the results of both methods comparable, we also compute the credible intervals from the profile likelihood.

DOF/CL	68%	95%
1	0.989	3.841
2	2.279	5.991

Table 3.1: χ^2 values to set the 68% and 95%CL limits on a distribution with 1 or 2 DOF.

This section discusses SFitter as tool of choice, by discussing the construction of an exclusive likelihood, the corresponding treatment of uncertainties and correlations, and different evaluation treatments, we can put it into action now. Based on the EFT-framework for SMEFT and the description of EDMs via a weak-scale EFT, introduced in Ch. 2, we perform the actual global analyses.

Chapter 4

Global Analysis using Published Likelihoods

The research and results presented in this chapter are based on work in collaboration with Maeve Madigan, Tilman Plehn, and Nikita Schmal and have been published in Ref. [1]. All figures and tables, as well as parts of the text, are similar or identical to the ones in the article.

Model-agnostic approaches are a promising avenue for understanding observed deviations from the SM. Thus, we choose SMEFT as a high-energy EFT expansion of the SM discussed in depth in Ch. 2. As model parameters we introduce the WCs of the SMEFT framework. To capture these deviations most efficiently, as much data as possible from different measurement types has to be included. Therefore, we perform a global analysis incorporating different ATLAS and CMS measurements using SFitter, which is explained in detail in Ch. 3. In particular we want to test the top sector [79, 121–130], since it provides a combination with the bottom sector [131–136] and a much broader set of precision measurements [137], which can then eventually test the impact of flavor symmetries. Recently, several groups also have performed a combined SMEFT analysis of the electroweak and top sectors [138,139], SMEFT analyses combined with parton density extraction [140–143], and even SMEFT analyses with lighter new particles [144–146].

The top sector provides unique aspects in the systematic searches for BSM physics that do not exist in the electroweak sector. From a theory perspective, the top sector is closely linked to the hierarchy problem or the dynamic origin of the Higgs vacuum expectation value, one of the problems that we are trying to improve our understanding with measurements from the LHC [147]. Experimentally, the precision measurements performed at the LHC go beyond simple kinematic distributions of top pair productions and extend to associated top pair production with gauge bosons, single top production, and top decay kinematics. This effort is beautifully matched by precision predictions [148]. Combining both aspects phenomenologically, the top groups from the ATLAS and CMS experiments provide experimental results in a way that makes the implementation in an external global analysis easy and provides optimal uncertainty treatment. This includes unfolded rate measurements, unfolded kinematic distributions, and most recently published likelihoods [149–151].

The publication of experimental data in the format of public likelihoods is a major step in the way experimental results can be re-interpreted [152–157]. The used `HistFactory` format [158] allows for efficient use of likelihoods using software such as `pyhf` [159, 160], `Spey` [161], and the simplified likelihood frameworks [162]. In the classic BSM sector, many likelihoods have already been made public and analyzed [163]. In the top sector, public ATLAS likelihoods [149–151] should be used outside the collaborations. We aim

to fill this gap by using them as the basis of a global SMEFT analysis of the top sector using SFitter.

In this chapter, we provide updated results based on an earlier dimension-six SMEFT analysis of the top sector using the SFitter framework [79]. With its adaptive nature to include new measurements, models, and different formats, SFitter is a perfect tool for including public likelihoods in a global analysis and determining their impact. We will start by discussing the different types of data we included and the implementation of three public likelihoods in Sec. 4.2. We will then include these likelihoods in the first SFitter analysis of the electroweak and top sectors in Sec. 4.3. While the physics behind combining these two sectors is largely understood [138, 139], in our global SFitter analysis, we will focus on the impact of theory uncertainties. In addition, we will probe the impact of a profile likelihood compared to Bayesian marginalization when extracting limits on single Wilson coefficients, where we saw significant effects on the Higgs and electroweak sector [75].

4.1 Setup

In this section, we will discuss the setup and implementation of public likelihoods in SFitter. A detailed explanation of the SMEFT sector and all of the Wilson coefficients used in this global analysis is provided in Ch. 2. Therefore, we only briefly discuss the SMEFT Lagrangian and used Wilson coefficients here and directly start by discussing the used data and prediction formats and their implementation in SFitter.

4.1.1 Data, predictions, and uncertainties

Predictions in the top sector are governed by the effective Lagrangian shown in Eq. (2.10) and truncated beyond mass dimension six. Additionally, we assume the separate $U(2)$ symmetries from Eq. (2.11) for the first and second generations. Next, the Wilson coefficients implemented in the top sector are discussed in Eqs. (2.12)-(2.14). The effect of the different operators and WCs on the different LHC observables are summarized in Tab. 4.1. Here, the main question is which operators modify the LHC rate and kinematic predictions through interference with the SM-matrix element, which only contributes at dimension-6 squared order. Next, we will elaborate further on implementing public likelihoods and how they are constructed and provided.

With the combined SMEFT analysis of the Higgs, electroweak, and top sector as a physical goal, the technical goal of this study is to integrate published experimental likelihoods into an SMEFT analysis for the first time. To simplify the procedure of including future experimental results as extensions to our current dataset, we analyze three measurements with publicly available likelihoods in the `HistFactory` [158] format on `HEPData`: An ATLAS measurement of the total inclusive $t\bar{t}$ cross section [149], an ATLAS measurement of the total inclusive $t\bar{t}Z$ cross section [150] and an ATLAS measurement of the total inclusive single-top cross section in the s-channel [151]. The implementation of these likelihoods into the SFitter framework will be discussed in more detail in Section 4.2. To obtain a realistic assessment of the effect of these likelihoods on the SMEFT, we incorporate them into a global analysis.

With this in mind, our analysis in the top sector considers all measurements listed in Tables 4.2 and 4.3. The analysis is an update to a previous global top analysis performed

	Wilson coeff.	$t\bar{t}$	single t	tW	tZ	t-decay	$t\bar{t}Z$	$t\bar{t}W$
Eq. (2.12)	$C_{Qq}^{1,8}$	Λ^{-2}	—	—	—	—	Λ^{-2}	Λ^{-2}
	$C_{Qq}^{3,8}$	Λ^{-2}	Λ^{-4} $[\Lambda^{-2}]$	—	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-2}	Λ^{-2}
	C_{tu}^{8}, C_{td}^{8}	Λ^{-2}	—	—	—	—	Λ^{-2}	—
	$C_{Qq}^{1,1}$	Λ^{-4} $[\Lambda^{-2}]$	—	—	—	—	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-4} $[\Lambda^{-2}]$
	$C_{Qq}^{3,1}$	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-2}	—	Λ^{-2}	Λ^{-2}	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-4} $[\Lambda^{-2}]$
	C_{tu}^{1}, C_{td}^{1}	Λ^{-4} $[\Lambda^{-2}]$	—	—	—	—	Λ^{-4} $[\Lambda^{-2}]$	—
Eq. (2.13)	C_{Qu}^{8}, C_{Qd}^{8}	Λ^{-2}	—	—	—	—	Λ^{-2}	—
	C_{tq}^{8}	Λ^{-2}	—	—	—	—	Λ^{-2}	Λ^{-2}
	C_{Qu}^{1}, C_{Qd}^{1}	Λ^{-4} $[\Lambda^{-2}]$	—	—	—	—	Λ^{-4} $[\Lambda^{-2}]$	—
	C_{tq}^{1}	Λ^{-4} $[\Lambda^{-2}]$	—	—	—	—	Λ^{-4} $[\Lambda^{-2}]$	Λ^{-4} $[\Lambda^{-2}]$
Eq. (2.14)	$C_{\phi Q}^{-}$	—	—	—	Λ^{-2}	—	Λ^{-2}	—
	$C_{\phi Q}^{3}$	—	Λ^{-2}	Λ^{-2}	Λ^{-2}	Λ^{-2}	Λ^{-2}	—
	$C_{\phi t}$	—	—	—	Λ^{-2}	—	Λ^{-2}	—
	$C_{\phi tb}$	—	Λ^{-4}	Λ^{-4}	Λ^{-4}	Λ^{-4}	—	—
	C_{tZ}	—	—	—	Λ^{-2}	—	Λ^{-2}	—
	C_{tW}	—	Λ^{-2}	Λ^{-2}	Λ^{-2}	Λ^{-2}	—	—
	C_{bW}	—	Λ^{-4}	Λ^{-4}	Λ^{-4}	Λ^{-4}	—	—
	C_{tG}	Λ^{-2}	$[\Lambda^{-2}]$	Λ^{-2}	—	$[\Lambda^{-2}]$	Λ^{-2}	Λ^{-2}

Table 4.1: Wilson coefficients and their contributions to top observables via SM-interference (Λ^{-2}) and via dimension-6 squared terms only (Λ^{-4}). A square bracket indicates that the Wilson coefficient contributes to the interference at NLO in QCD. Table adapted from Ref. [75].

by SFitter in Ref. [79]. In Tables 4.2 and 4.3, the measurements that are new relative to those included in Ref. [79] are highlighted, as well as those for which a public likelihood is available. Whenever possible, we prefer to include measurements containing the full Run II luminosity and choose such measurements that are in the boosted regime in which the sensitivity to energy-growing SMEFT operators is maximized; see, for example, the top pair production invariant mass distribution of Ref. [179]. Thus, the dataset consists of a total of 122 data points spanning the $t\bar{t}$, $t\bar{t}+X(Z,W,\gamma)$ and single top (s, t-channel, tW and tZ) sectors, including measurements of top-pair production charge asymmetries A_C and W boson polarization in top decays (F_0, F_L).

Precision predictions from perturbative quantum field theories are a key ingredient of global analyses to obtain a good baseline. Thus, most observables in this analysis unfold at the parton level while assuming stable top quarks. This allows us to use fixed-order calculations to determine the SM predictions at NLO in QCD using `MadGraph5 aMC@NLO` [217, 218] and `NNPDF 4.0` [219] interfaced with `LHAPDF` [220]. Alongside the observables listed in Tables 4.2 and 4.3, we note whether the SM predictions for these observables are approximated at NNLO in QCD using a K-factor approximation and referencing the source of these QCD k-factors. In the case of new top quark pair production observables, these QCD k-factors are calculated using `HighTea` [148].

For calculations at NLO in QCD on the effect of the SMEFT on all updated measurements in the top sector, we used the FeynRules [221] model `SMEFTatNLO` [222] up to quadratic order in the EFT expansion. The exceptions made are the measurements of the $t\bar{t}\gamma$

total cross sections at 8 TeV by ATLAS [192] and CMS [191], for which the SMEFT predictions at LO in QCD are taken from Ref. [141].

We must also consider theory uncertainties, which appear when comparing a measurement to a first-principle description. They might cover a wide range of approximations that can be calculated, such as an LHC cross section obtained from a fundamental renormalized Lagrangian. The truncation of the perturbative series in QCD and the electroweak gauge coupling dominates the LHC. Because these perturbative series converge very slowly for LHC rates, theory predictions have become limiting factors for interpreting many LHC measurements in terms of actual physics. Aside from the size of the theory uncertainties,

Experiment	E [GeV]	$\mathcal{L}$ [fb^{-1}]	Channel	Obs.	# Bins	New	Likeli.	QCD k-factor
CMS [164]	8	19.7	$e\mu$	$\sigma_{t\bar{t}}$				[165]
ATLAS [166]	8	20.2	lj	$\sigma_{t\bar{t}}$				[165]
CMS [167]	13	137	lj	$\sigma_{t\bar{t}}$		✓		[165]
CMS [168]	13	35.9	ll	$\sigma_{t\bar{t}}$				[165]
ATLAS [169]	13	36.1	ll	$\sigma_{t\bar{t}}$		✓		[165]
ATLAS [170]	13	36.1	aj	$\sigma_{t\bar{t}}$		✓		[165]
ATLAS [149]	13	139	lj	$\sigma_{t\bar{t}}$		✓	✓	[165]
CMS [171]	13.6	1.21	ll, lj	$\sigma_{t\bar{t}}$		✓		[171]
CMS [172]	8	19.7	lj	$\frac{1}{\sigma}\frac{d\sigma}{dp_T^t}$	7			[173–175]
CMS [172]	8	19.7	ll	$\frac{1}{\sigma}\frac{d\sigma}{dp_T^t}$	5			[173–175]
ATLAS [176]	8	20.3	lj	$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	7			[173–175]
CMS [167]	13	137	lj	$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	15	✓		[148]
CMS [177]	13	35.9	ll	$\frac{1}{\sigma}\frac{d\sigma}{d\Delta y_{t\bar{t}}}$	8			[173–175]
ATLAS [178]	13	36	lj	$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	9	✓		[148]
ATLAS [179]	13	139	aj, high-p_T	$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	13	✓		
CMS [180]	8	19.7	lj	A_C				[181]
CMS [182]	8	19.5	ll	A_C				[181]
ATLAS [183]	8	20.3	lj	A_C				[181]
ATLAS [184]	8	20.3	ll	A_C				[181]
CMS [185]	13	138	lj	A_C		✓		[181]
ATLAS [186]	13	139	lj	A_C		✓		[181]
ATLAS [150]	13	139		$\sigma_{t\bar{t}Z}$		✓	✓	[187]
CMS [188]	13	77.5		$\sigma_{t\bar{t}Z}$				[187]
CMS [189]	13	35.9		$\sigma_{t\bar{t}W}$				[187]
ATLAS [190]	13	36.1		$\sigma_{t\bar{t}W}$		✓		[187]
CMS [191]	8	19.7		$\sigma_{t\bar{t}\gamma}$		✓		
ATLAS [192]	8	20.2		$\sigma_{t\bar{t}\gamma}$		✓		

Table 4.2: Top pair observables included in our global analysis. "New" is defined relative to the previous SFitter analysis [79]. "Likeli." indicates a dataset for which a public likelihood is available — further details of these datasets are provided in Sec. 4.2.

it is problematic that they do not follow any statistical pattern or model [223], and assuming a Gaussian distribution is neither justified nor conservative.

Because of their impact on global analyses of effective Lagrangians, SFitter puts an emphasis on the proper description of these uncertainties, including their correlations between different observables, as described in Ch. 3. In the top sector, the theory uncertainties are critical for the precisely measured top pair production rates [79] and are correlated between different final states for rate measurements. We typically use the theory uncertainties reported in the respective publications, with the exception that we enforce a minimum scale uncertainty of 10% for total rates in single top production and 2% for bin-wise kinematic distributions.

Experiment	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb^{-1}]	Channel	Obs.	# Bins	New	Likeli.	QCD k-factor
ATLAS [193]	7	4.59	t-ch	$\sigma_{tq+\bar{t}q}$				
CMS [194]	7	1.17 (e) 1.56 (μ)	t-ch	$\sigma_{tq+\bar{t}q}$				
ATLAS [195]	8	20.2	t-ch	σ_{tq}, $\sigma_{\bar{t}q}$				
CMS [196]	8	19.7	t-ch	σ_{tq}, $\sigma_{\bar{t}q}$				
ATLAS [197]	13	3.2	t-ch	σ_{tq}, $\sigma_{\bar{t}q}$				[198]
CMS [199]	13	2.2	t-ch	σ_{tq}, $\sigma_{\bar{t}q}$				[198]
CMS [200]	13	35.9	t-ch	$\frac{1}{\sigma}\frac{d\sigma}{d\lvert p_{T,t}\rvert}$	5	✓		
CMS [201]	7	5.1	s-ch	$\sigma_{t\bar{b}+\bar{t}b}$				
CMS [201]	8	19.7	s-ch	$\sigma_{t\bar{b}+\bar{t}b}$				
ATLAS [202]	8	20.3	s-ch	$\sigma_{t\bar{b}+\bar{t}b}$				
ATLAS [151]	13	139	s-ch	$\sigma_{t\bar{b}+\bar{t}b}$		✓	✓	
ATLAS [203]	7	2.05	tW ($2l$)	$\sigma_{tW+\bar{t}W}$				
CMS [204]	7	4.9	tW ($2l$)	$\sigma_{tW+\bar{t}W}$				
ATLAS [205]	8	20.3	tW ($2l$)	$\sigma_{tW+\bar{t}W}$				
ATLAS [206]	8	20.2	tW ($1l$)	$\sigma_{tW+\bar{t}W}$		✓		
CMS [207]	8	12.2	tW ($2l$)	$\sigma_{tW+\bar{t}W}$				
ATLAS [208]	13	3.2	tW ($1l$)	$\sigma_{tW+\bar{t}W}$				
CMS [209]	13	35.9	tW ($e\mu j$)	$\sigma_{tW+\bar{t}W}$				
CMS [210]	13	36	tW ($2l$)	$\sigma_{tW+\bar{t}W}$		✓		
ATLAS [211]	13	36.1	tZ	σ_{tZq}				
ATLAS [212]	7	1.04		F_0, F_L				
CMS [213]	7	5		F_0, F_L				
ATLAS [214]	8	20.2		F_0, F_L				
CMS [215]	8	19.8		F_0, F_L				
ATLAS [216]	13	139		F_0, F_L		✓		

Table 4.3: Single top and top decay observables included in our global analysis. "New" is defined relative to the previous SFitter analysis [79]. "Likeli." indicates a dataset for which a public likelihood is available — further details of these datasets are provided in Sec. 4.2.

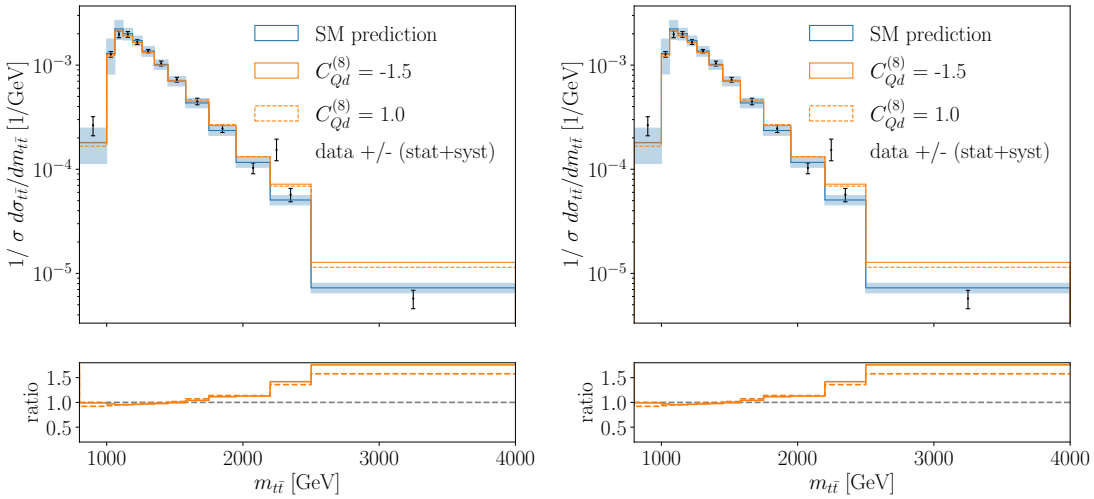

Figure 4.1: Left: Impact of $\mathcal{O}_{Qd}^{(8)}$ on the unfolded ATLAS $m_{t\bar{t}}$ distribution in the lepton+jets channel [178]. Right: Impact of this operator on the unfolded ATLAS $m_{t\bar{t}}$ distribution in the all-hadronic channel measured with boosted top quarks [179].

Boosted top pair production

As part of our dataset, we highlight the reinterpretation of the ATLAS measurement of $t\bar{t}$ production in the lepton+jets channel [178] and the ATLAS measurement of $t\bar{t}$ production using boosted top quarks in the all-hadronic channel [179]. Both are differential in the top-pair invariant mass, as shown in Fig.4.1. The measurement using boosted top quarks is unfolded to a fiducial Parton-level phase space, defined by

$$p_{T,t_1} > 500 \text{ GeV} \qquad \text{and} \qquad p_{T,t_2} > 350 \text{ GeV} , \tag{4.1}$$

allowing for an easy comparison with fixed-order calculations. This, alongside the high-$m_{t\bar{t}}$ reach up to 4 TeV of this distribution, makes it an excellent candidate for constraining the energy-growing SMEFT four-fermion operators of the top sector. We display the impact of one of these operators, $\mathcal{O}_{Qd}^{8}$, in Fig.4.1.

The theory uncertainty is shown in blue in both figures and compared to the statistical and systematic uncertainties in the experimental data. In both cases, the values of $\mathcal{C}_{Qd}^{8}$ chosen are those which would produce a 3σ deviation in a one-parameter analysis. We observe that, while the measurement unfolded to the full phase space is sensitive to the energy-growing effects of $\mathcal{O}_{Qd}^{8}$, this sensitivity is significantly enhanced by the measurement of boosted top quarks. The lower panel in both figures shows the ratio of the upper and lower bounds of $\mathcal{C}_{Qd}^{8}$ and the SM prediction.

4.1.2 SFitter as framework

The SFitter framework, as in depth discussed in Ch. 3, is a tool for global analyses of LHC measurements in the context of BSM physics. There, the relation of this data to full models with a proper uncertainty treatment in precision matching is crucial because the typical scale variation between directly probed energies and those that are

indirectly accessible is not very large. On the other hand, the assumed EFT description is not universal but only defined by possible on-shell propagators in the observables and relative to the UV completion and its typical coupling strengths. Without additional information on the underlying model, the Lagrangian in Eq.(2.10) is degenerate along $C_k \sim \Lambda^2$, which means the EFT assumption of large Λ improves for larger postulated couplings. This is the reason why SFitter SMEFT analyses start with the truncated dimension-6 Lagrangian at face value. Tab. 4.1 shows that some WCs only enter with dimension-6 squared contributions to the LHC observables. Therefore, we truncate the Lagrangian rather than the LHC rate prediction. All of these assumptions always need to be validated for a given dataset and UV-Completion [102, 224–226]. However, the assumptions we made in this global analysis ensure that we optimally use all kinematic information provided, especially those in the tail regions of the considered distributions.

As mentioned earlier, the heart of SFitter is the extraction of a fully exclusive likelihood for a given measurement d from Sec. 4.1.1, evaluated over a combined Wilson coefficient c space with nuisance parameters θ,

$$p(d|c,\theta) = \text{Pois}(d|m(c,\theta,b))\ \text{Pois}(b_{\text{CR}}|b\,k)\ \prod_i \mathcal{C}_i(\theta_i,\sigma_i)\ . \tag{4.2}$$

This likelihood also incorporates the effects of statistical, systematic, and theory uncertainties. The first Poisson distribution gives the probability to observe d events given the corresponding theory prediction $m(c,\theta,b)$, which in turn depends on the predicted background count b. The background rate is, itself, constrained by measurements b_{CR} in the control region, implemented as a scaled prediction kb with a suitable factor k. The constraint function $\mathcal{C}$ gives the distribution of the nuisance parameter θ_i, given a width measure σ_i, depending on the source of uncertainty. As discussed in Ch. 3, the distribution can either follow a Gaussian, Poissonian, or flat distribution.

The flat scale uncertainty, which is considered a theory uncertainty, is not parametrization invariant, as one would expect from a fixed range, but without a preferred central value, we consider it conservative. They are obtained in the top sector by varying the renormalization and factorization scales μ_R and μ_F by a factor of 2 around their respective central value. These are process dependent and chosen to be $\mu_R = \mu_F = m_t + \frac{1}{2}m_V$ for associated $t\bar{t}$ production with $V = W, Z$. For $t\bar{t}$ production, the sum of the transverse masses of the top and anti-top is used, while for single top production, they are set to the top mass m_t.

By ansatz, we construct all measurements d in SFitter are uncorrelated, and it constructs an individual likelihood for each measurement, as defined in Eq (4.2). With this, the full likelihood is constructed as a product of individual contributions. As pointed out in Ch. 3, we include systematic uncertainties as correlated Gaussians, while theory uncertainties are correlated for all measurements with identical predictions. They are also correlated within one measurement across all bins but not across several different measurements.

In the SMEFT sector for constructing the exclusive likelihood, SFitter uses cross section predictions over the entire model parameter space and extracts the quadratic behavior analytically, which guarantees sufficient precision even for small Wilson coefficients. Then, an MCMC algorithm is used to evaluate the likelihood in Eq. (4.2) numerically. Finally, to combine uncertainties by removing nuisance parameters or to reduce the space of physical Wilson coefficients, SFitter can employ a profile likelihood or a Bayesian marginalization [75, 99]. Obviously, these two methods give different results. Only for uncorrelated Gaussians do the profile likelihood and Bayesian marginalization lead to

the common result of errors added in quadrature. For a flat likelihood, the uncorrelated profile likelihood adds the two uncertainties linearly, which happens for the scale and PDF uncertainty in SFitter. The profile likelihood combination of a flat and a Gaussian uncertainty gives the well-known RFit prescription [227]. In contrast, when applying marginalization on the combination of Gaussian and flat uncertainties, the central limit theorem ensures that the final posterior will be Gaussian again.

4.2 Public likelihoods

For a standard SFitter analysis, we extract systematic uncertainties for each measurement from the respective experimental publications. Systematics of the same type are fully correlated between measurements of the same experiment. This approach has drawbacks. For instance, we can only use the uncertainty categories reported in the experimental publications or on `HEPData`, and this information often needs to be extracted by hand. Public likelihoods include the full information on many systematic uncertainties in a documented manner, making their implementation more accurate and efficient.

Likelihoods are published in the `HistFactory` format [158], similar to the SFitter likelihood in Eq.(4.2). For each bin b measured as a kinematic distribution of a given channel or final state, it provides

$$p(d_b|\mu,\theta) = \text{Pois}(d_b|m_b(\mu,\theta)) \prod_i \mathcal{C}_i(a_i|\theta_i) \,, \tag{4.3}$$

where d_b and m_b are the measured and expected number of events in bin b. The nuisance parameters θ_i are constrained by $\mathcal{C}_i(a_i|\theta_i)$ with the auxiliary data a_i. The parameter of interest μ describes, for instance, a signal strength. This corresponds to the Wilson coefficients shown in Eq.(4.2).

The analysis of these publicly provided likelihoods is performed using `pyhf` [159, 160], a python module allowing for easy construction of `HistFactory` likelihoods and their subsequent statistical analysis. It uses data published in the JSON format to compute

Description	Modification	Constraint $\mathcal{C}$
Luminosity ("lumi")	$\kappa_{sb} = \lambda$	$\mathcal{N}(l = \lambda_0\|\lambda, \sigma_\lambda)$
Normalization unc. ("normsys")	$\kappa_{sb} = g_p(\alpha\|\kappa_{sb,\alpha=\pm1})$	$\mathcal{N}(a = 0\|\alpha, \sigma = 1)$
Correlated Shape ("histosys")	$\Delta_{sb} = f_p(\alpha\|\Delta_{sb,\alpha=\pm1})$	$\mathcal{N}(a = 0\|\alpha, \sigma = 1)$
MC Stat. ("staterror")	$\kappa_{sb} = \gamma_b$	$\prod_b \mathcal{N}(a_{\gamma_b} = 1\|\gamma_b, \delta_b)$
Uncorrelated Shape ("shapesys")	$\kappa_{sb} = \gamma_b$	$\prod_b \text{Pois}(\sigma_b^{-2}\|\sigma_b^{-2}\gamma_b)$
Normalization ("normfactor")	$\kappa_{sb} = \mu$	

Table 4.4: List of modifiers in the construction of the `HistFactory` likelihoods, adapted from Ref. [228]. Per-bin modifiers are denoted as γ_b, while interpolated modifiers are denoted as α. Here g_p and f_p describe different interpolation strategies used to compute these from the values $\kappa_{sb,\alpha=\pm1}, \Delta_{sb,\alpha=\pm1}$ provided in the likelihood. Luminosity and scale factors affect all bins equally and are denoted as λ and μ, respectively.

the predicted number of events using

$$m_b = \sum_s \left(\prod_\kappa \kappa_{sb} \right) \left(\bar{m}_{sb} + \sum_\Delta \Delta_{sb} \right) , \tag{4.4}$$

with the nominal expected rate $\bar{m}_{sb}$ and multiplicative (κ_{sb}) and additive (Δ_{sb}) modifiers for each physics process s. These modifiers correspond to the nuisance parameters affecting the event rate m_b. The type of modifier and the constraints on its corresponding nuisance parameter depend on the uncertainty type. The most common ones are given in Tab. 4.4. Using the public likelihoods in terms of modifiers and nominal rates $\bar{m}_{sb}$, we can reproduce the experimental results. For visualization, we use `cabinetry` [229], a python library making use of `pyhf` for statistical analyses.

Starting from a public likelihood, we can organize the full set of nuisance parameters with respect to their systematic uncertainties in a few categories. This allows for an easier numerical treatment at no cost. To compute the ranges of the nuisance parameters for these categories, we start by using a profile likelihood to determine the corresponding central values, and as a second step, we perform an analysis for the distribution of the nuisance parameter. In this section, we will show the implementation and corresponding testing of three public ATLAS likelihoods.

4.2.1 ATLAS $t\bar{t}$ likelihood

As the first public likelihood, we analyze the $t\bar{t}$ rate measurement in the lepton+jets final state [149]. The likelihood consists of three channels or signal regions, using the aplanarity, minimum lepton-jet mass, and average angular distance between jets. In this case, the parameter of interest μ is in the $t\bar{t}$ signal strength, with a total of 177 nuisance parameters covering the systematic uncertainties.

To test our implementation and evaluation of the public likelihood, we first reproduce some key results from the ATLAS analysis in Fig. 4.2. Therefore, we show the values for

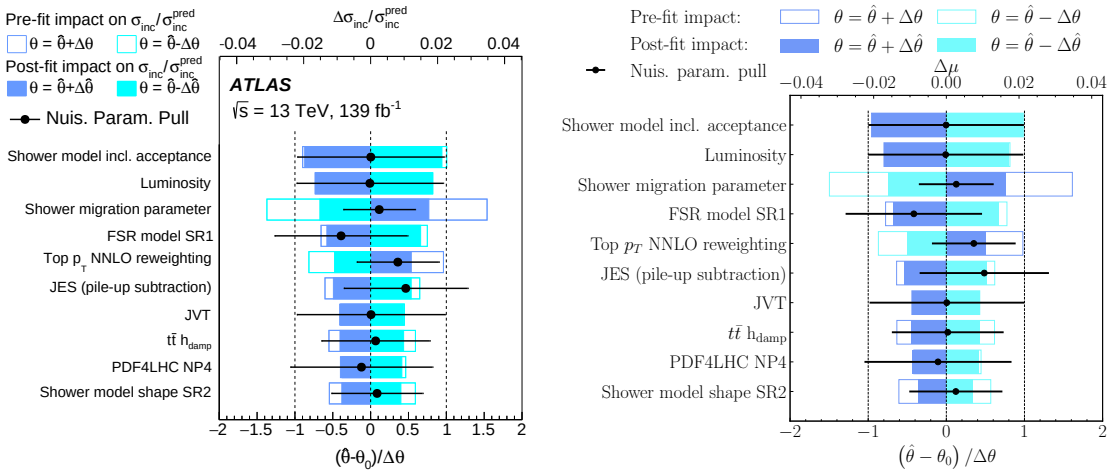

Figure 4.2: Impact of nuisance parameters on the $t\bar{t}$ total rate fit. We compare the ATLAS result [149] (left) and our evaluation of the public likelihood (right).

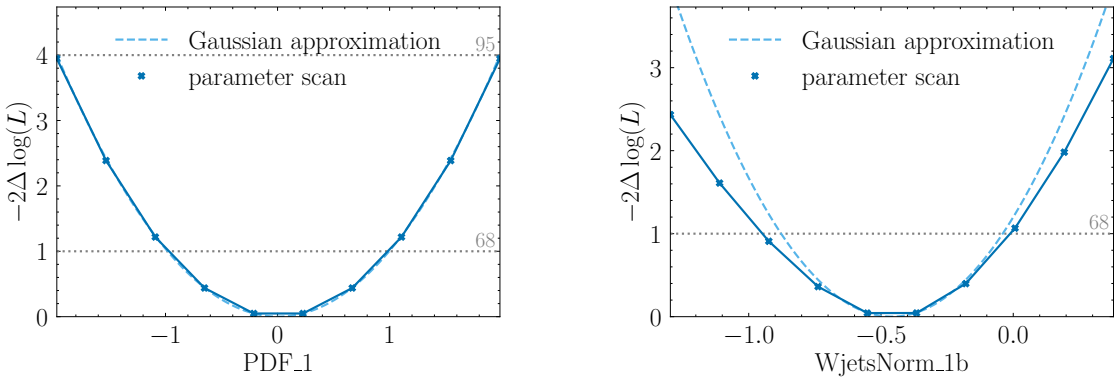

Figure 4.3: Dependence of the log-likelihood for the $t\bar{t}$ rate on two nuisance parameters, one describing the PDF uncertainty and one describing the $W+$ jets background normalization uncertainty, compared with the Gaussian approximation.

the single nuisance parameters that maximize the likelihood and corresponding pulls,

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \,. \tag{4.5}$$

Here, $\hat{\theta}$ describes the maximum likelihood values, and θ_0 is the value before the fit, normalized to the pre-fit uncertainty $\Delta\theta$. We also show the impact of the individual nuisance parameters on the signal strength μ. This impact is determined by repeating the fit after fixing the nuisance parameter to its maximum-likelihood value $\hat{\theta}$, shifted by its prefit (postfit) uncertainties $\pm\Delta\theta(\pm\Delta\hat{\theta})$. The left panel of Fig. 4.2 is taken from Ref. [149], while the right panel shows our reproduced results. Both sets show excellent agreement, with negligible differences for a few select nuisance parameters.

As a next step, we analyze the full likelihood as a function of a single nuisance parameter.

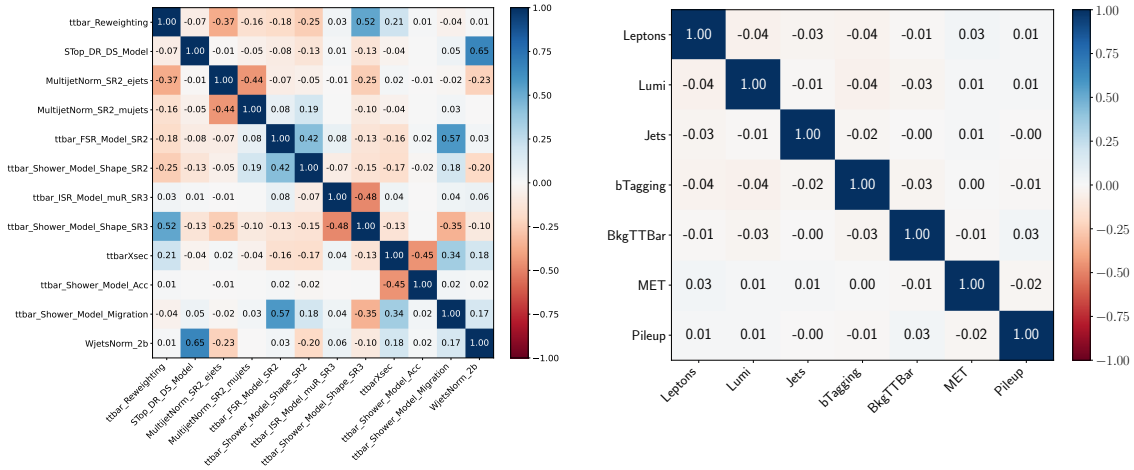

Figure 4.4: Left: Correlations between individual nuisance parameters affecting the $t\bar{t}$ rate with at least one correlation greater than 0.4. Right: Correlations between categories of systematic uncertainties extracted from the $t\bar{t}$ likelihood as implemented in SFitter.

This allows us to check the validity of the assumption of a Gaussian likelihood, which is assumed for systematic uncertainties in SFitter. For each nuisance parameter, we generally find excellent agreement with the Gaussian assumption, as shown on the left-hand side, with only a few exceptions. In Fig. 4.3, one parameter shows excellent agreement and one parameter shows poor agreement. We are even able to control the larger deviations, showing a good agreement with the Gaussian approximation when we translate them into terms of standard deviations. The categorization of nuisance parameters washes out non-Gaussian shapes in these exceptions when categorized with Gaussian-shaped parameters.

Finally, we test the correlations between individual and categorized nuisance parameters; for later, we use the categories implemented in SFitter. The left panel of Fig. 4.4 shows the correlations of all individual nuisance parameters with at least one correlation greater than 0.4. Since the public likelihoods do not provide additional metadata on all nuisance parameters, their labels do not necessarily match those used in the impact plots. We find that out of the many nuisance parameters included in the public likelihood, only very few are significantly correlated. The strong correlations are mainly seen for nuisance parameters impacted by modeling choices or jets.

In the standard SFitter profiling approach, we group these individual nuisance parameters into uncorrelated categories and implement these categories with a single nuisance parameter each. Standard categories cover leptons, jets, tagging, and luminosity. Additional categories are process-specific, such as certain backgrounds or missing transverse energy. For all processes in our dataset, we use 21 nuisance parameters describing the systematic uncertainties assumed to be uncorrelated. Using the full likelihood, we show

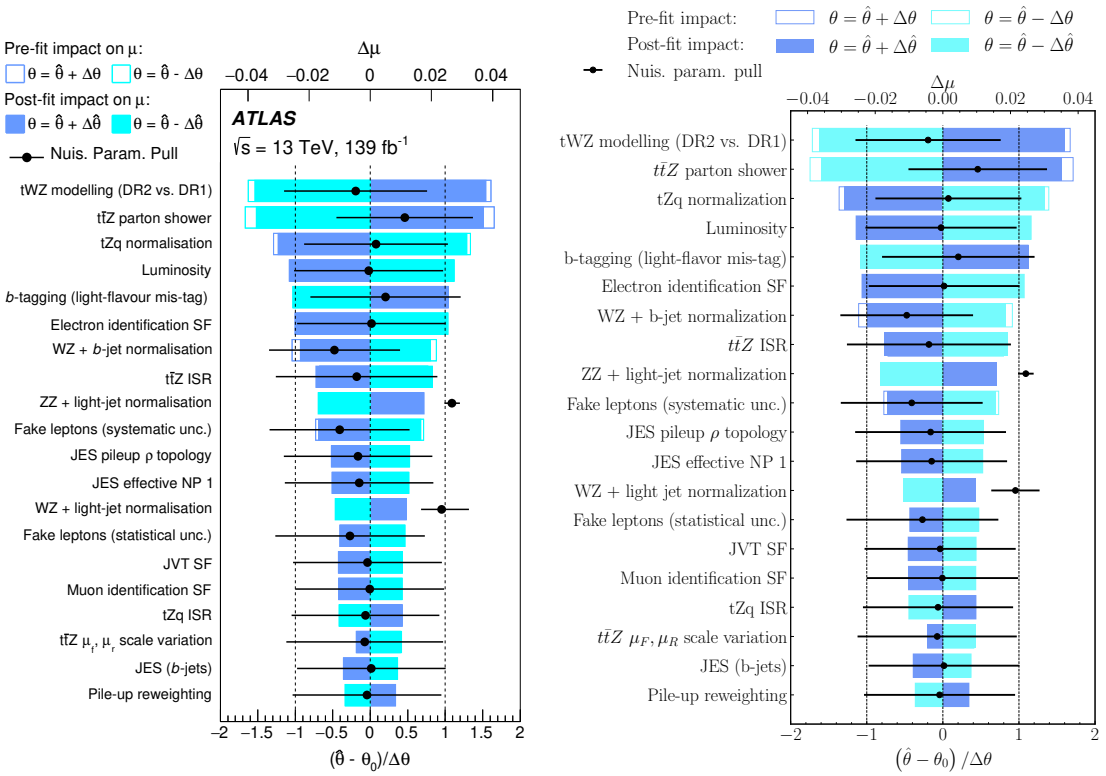

Figure 4.5: Impact of nuisance parameters on the $t\bar{t}Z$ total rate fit. We compare the ATLAS result [150] (left) and our evaluation of the public likelihood (right).

the correlations between these categories in the right panel of Fig. 4.4. The fact that the correlations between categories essentially vanish validates this SFitter approach.

4.2.2 ATLAS $t\bar{t}Z$ likelihood

The second likelihood we implement is from the $t\bar{t}Z$ rate measurement [150]. It simultaneously fits both 3-lepton and 4-lepton signal regions and the corresponding control regions. The parameter of interest is the $t\bar{t}Z$ signal strength. A total of 230 nuisance parameters are provided to describe the systematic uncertainties. Unlike for the $t\bar{t}$ likelihood, there are no uncertainties on the shape of the signal since each signal region is described by a single bin.

Following the method described for the $t\bar{t}$ analysis, we also test the $t\bar{t}Z$ likelihood and our implementation into SFitter. Figure 4.5 compares the impact of the nuisance parameters and pulls taken from Ref. [150] with those reproduced by us. We see excellent agreement for all nuisance parameters.

Same as for the previous likelihood implementation, we show the correlations between different nuisance parameters, with at least one correlation greater than 0.3 in the left panel of Fig. 4.6. Then, we can compare them to the SFitter implementation shown in the right panel. We find that the correlations between these individual nuisance parameters are already much smaller in this likelihood compared to the previous one. The only strong correlation seems to appear between scale uncertainties and the signal strength of the corresponding background. Consequently, the results, after combining all nuisance parameters into the SFitter categories, display negligible correlations between categories.

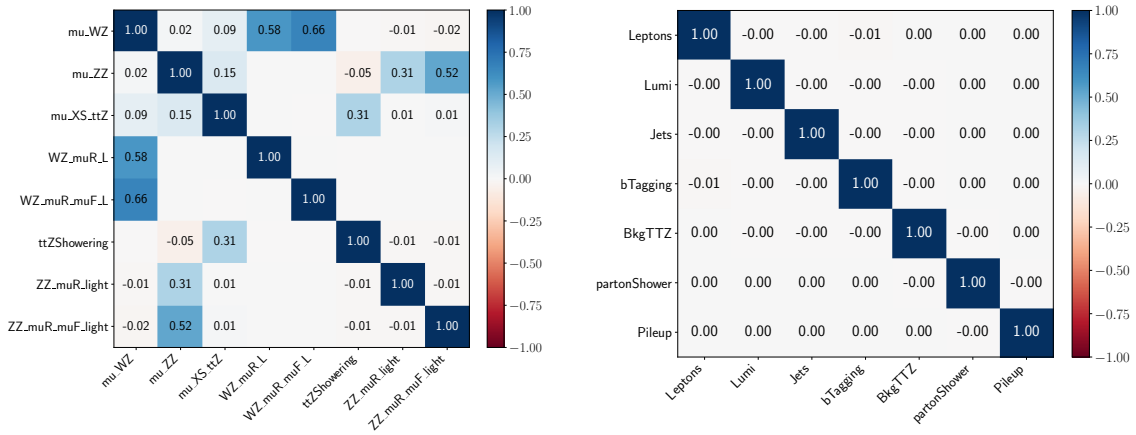

Figure 4.6: Left: Correlations between individual nuisance parameters affecting the $t\bar{t}Z$ rate with at least one correlation greater than 0.3. Right: Correlations between categories of systematic uncertainties extracted from the $t\bar{t}Z$ likelihood as implemented in SFitter.

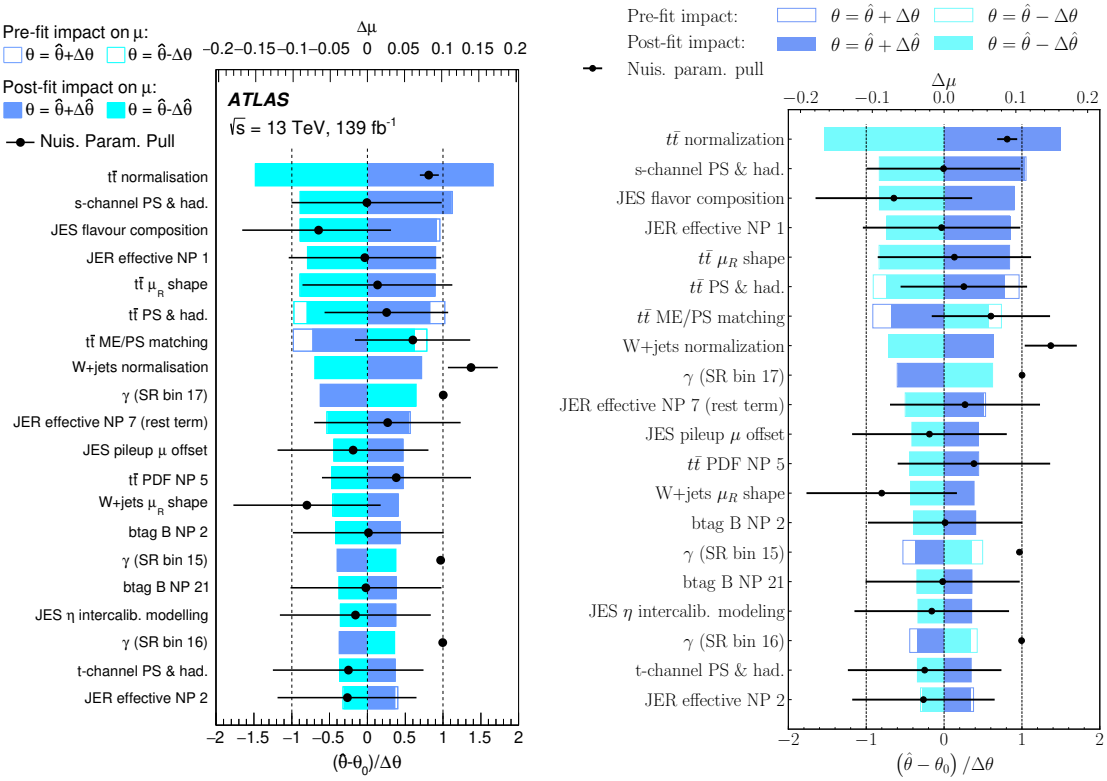

Figure 4.7: Impact of nuisance parameters on the s-channel single top rate fit. We compare the ATLAS result [151] (left) and our evaluation of the public likelihood (right).

4.2.3 ATLAS s-channel single top likelihood

Lastly, we implement the likelihood for a signal strength measurement of an s-channel single top production [151]. Unlike the previous measurements, it only consists of a single channel, using the matrix element method (MEM) to determine the probability that an event is a signal event. The discriminant defined using the MEM gives a distribution with 171 nuisance parameters affecting the rate and shape of the signal.

Once again, we validate our implementation of this likelihood in Fig. 4.7, showing the impact and pulls of different nuisance parameters from Ref. [151] in the left panel and our reproduction on the right. We find perfect agreement, which shows that regardless of the process considered, the public likelihoods allow for easy and precise reproduction of the experimental results in more detail than most global analyses will ever need or want to use.

The correlations in the left panel of Fig. 4.8 show strong correlations between select nuisance parameters. The strongest correlations appear between jet-related uncertainties and the signal strengths of the two dominant backgrounds, $t\bar{t}$ and W+jets. For SFitter, these nuisance parameters are put into the background uncertainty category. These strong correlations are therefore implicitly included in this larger category, and the final implementation into SFitter is essentially uncorrelated, as one can see in the right of Fig. 4.8. While one still finds a nonzero correlation between the jet and background uncertainties and between the jet and lepton uncertainties, these are all negligibly small.

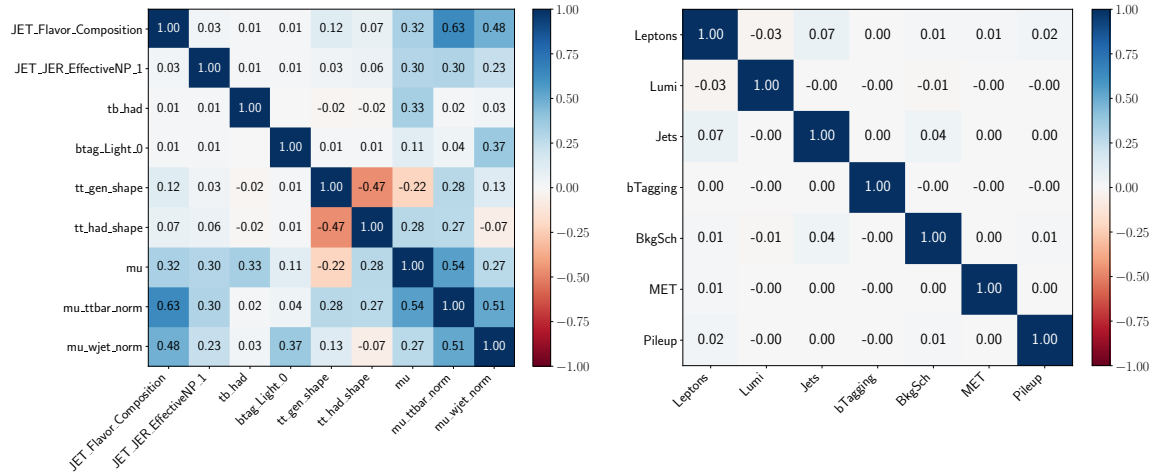

Figure 4.8: Left: Correlations between individual nuisance parameters affecting the s-channel single top rate with at least one correlation greater than 0.3. Right: Correlations between categories of systematic uncertainties extracted from the single top likelihood as implemented in SFitter.

4.3 Global analysis

Using, for the first time, public likelihoods in a global SMEFT analysis allows us to look at different relevant questions than in previous global analyses. From the data included in our analysis, we know that our global analysis is somewhat unlikely to uncover a fundamental and statistically significant breakdown of the SM. Therefore, we first look at the impact on the constraining power from new measurements, especially boosted top kinematics, relative to Ref. [79]. We then study the impact of correlated uncertainties encoded in the public likelihoods. From a pure statistics perspective, we will also check if lower-dimensional limits, which are either extracted by profiling or by marginalization, differ. Finally, we provide SMEFT limits combining the updated top sector analysis with the electroweak and Higgs sector from Ref. [75].

4.3.1 Better and boosted measurements

Before we study more conceptual questions of global SMEFT analyses, we update our dataset with new measurements, as marked in Tabs. 4.2 and 4.3. In Fig. 4.9, we show the constraints on a selection of 2-dimensional correlations of Wilson coefficients, using all top data compared to the previous SFitter top analysis [79]. All constraints are the result of an analysis of all 22 Wilson coefficients. To extract limits on pairs of coefficients, we use a profile likelihood approach. Potential differences in marginalization will be discussed in Sec. 4.3.3.

The left panel in Fig. 4.9 shows the impact on the four-fermion operators $\mathcal{O}_{Qu}^{1}$ and $\mathcal{O}_{tq}^{8}$. Both operators receive constraints from top pair production, now with a public likelihood [149], as well as new data in the boosted regime [179] and at 13.6 TeV [171]. The measurements, including boosted kinematics [179], provide the dominant constraining power and will be discussed in more detail below.

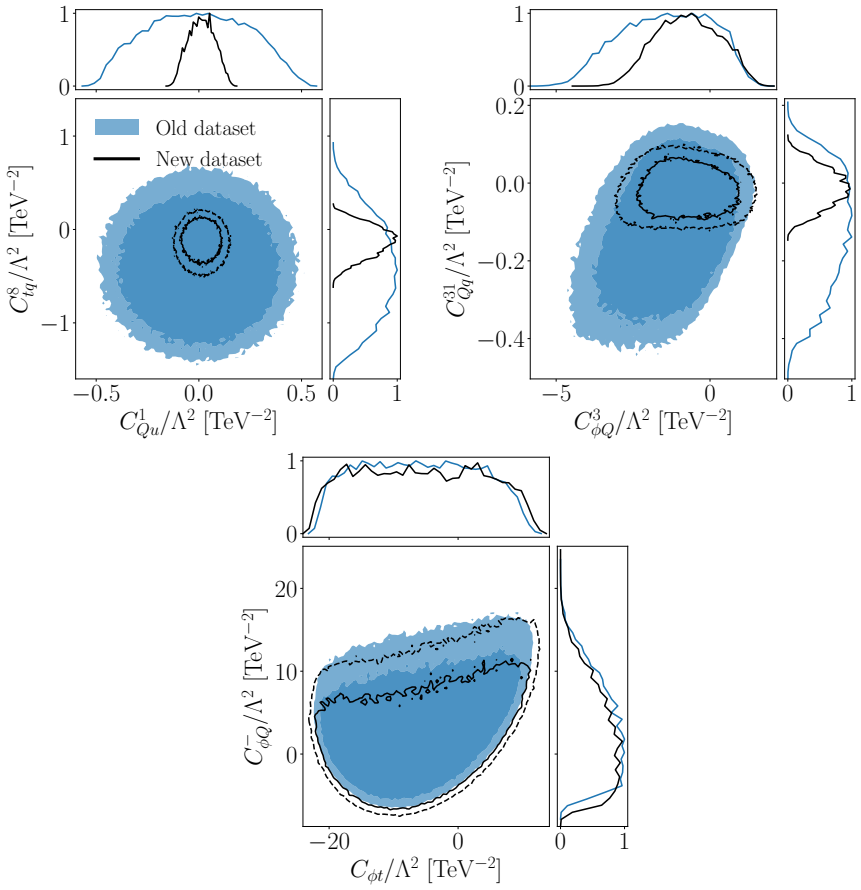

Figure 4.9: Profile likelihood correlations for three pairs out of the 22 Wilson coefficients, illustrating the impact of the new data listed in Tabs. 4.2 and 4.3 (black) compared to the previous analysis [79] (blue).

The right panel shows the improvement in constraints on $\mathcal{O}_{\phi Q}^{3}$ and $\mathcal{O}_{Qq}^{31}$. With single top production providing constraints on them, and we again benefit from the public likelihood [151], a new $p_{T,t}$ distribution in t-channel single top production [200], and new measurements of the tW production cross section [206, 210]. We observe an improvement in the individual constraints and their correlation. In particular, $\mathcal{O}_{Qq}^{31}$ receives some constraining power from boosted top pair production, which in turn allows single top measurements to constrain $\mathcal{O}_{\phi Q}^{3}$.

Finally, in the lower panel of Fig. 4.9 we highlight the improvement in probing $\mathcal{O}_{\phi t}$ and $\mathcal{O}_{\phi Q}^{-}$. As before, these operators are constrained by measurements of $t\bar{t}Z$ production, for which we use a public likelihood. However, in this case, we only find a small change in the correlated likelihood.

Altogether, we find that the public likelihoods do not have a significant effect on our SMEFT limits. As discussed in Sec. 4.2, the likelihoods available and included in our analysis all describe total cross sections, which limits their impact. On the positive side, public likelihoods allow for accurate modeling of correlated systematics, an aspect we will discuss in Sec 4.3.2.

Much of the improvement we see from our new dataset is due to the boosted regime. For SMEFT analyses, such measurements are extremely helpful to constrain operators,

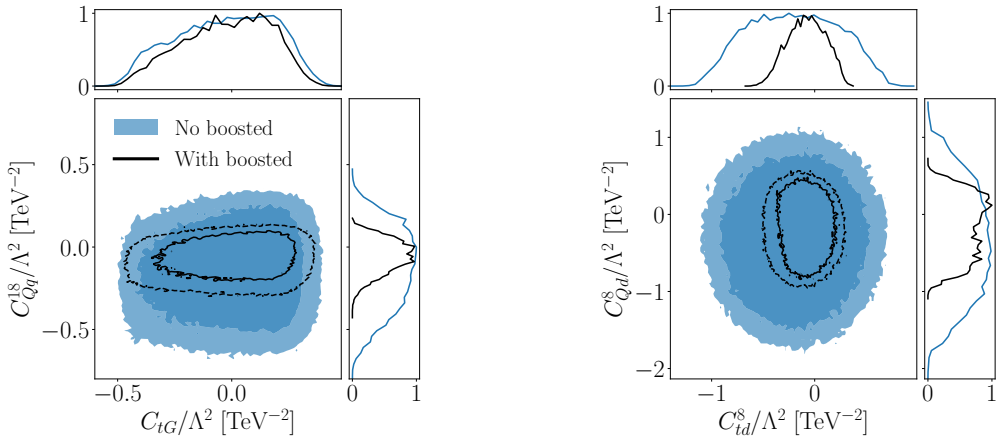

Figure 4.10: Profile likelihood correlations showing the impact of boosted top pair kinematics [179] (black), compared to the same dataset without this one measurement (blue).

including momentum scaling. As discussed in Sec. 4.1.1, we add the unfolded ATLAS measurement of boosted top pair production [179]. In Fig. 4.1, we already showed the impact of a single SMEFT operator $\mathcal{O}_{Qd}^{8}$ on the normalized $m_{t\bar{t}}$ distribution of this measurement.

Here, we study its effect on the global analysis. Figure 4.10 demonstrates the impact on a selection of two-operator correlations. The complete analysis including all data in Tab. 4.2 and Tab. 4.3 is compared to the case where the measurement of boosted tops from Ref. [179] is excluded. In the first panel, we observe an increase in constraining power on $\mathcal{O}_{Qq}^{18}$, while the constraints on $\mathcal{O}_{tG}$ are stable. This follows because this operator is instead constrained by the $t\bar{t}$ total cross section. In contrast, the limits on the Wilson coefficients for energy-growing 4-fermion operators improve by a factor of two, as shown in the right-hand panel for $\mathcal{O}_{Qq}^{8}$ and $\mathcal{O}_{td}^{8}$.

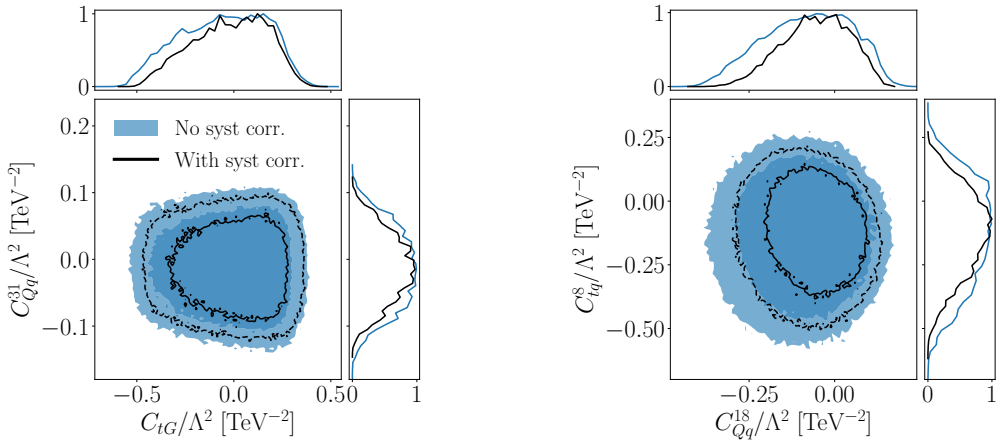

Figure 4.11: Profile likelihood correlations including correlated systematic and theory uncertainties (blue) versus ignoring correlations between experimental systematics (black).

4.3.2 Correlated systematics

Public likelihoods, as discussed in Sec. 4.2, allow us to model and study correlated systematic uncertainties across measurements by the same experiment. For measurements without public likelihoods, the approximate treatment of correlations is discussed in Sec. 4.1.2. For the Higgs sector, we already know that the correlations of systematic uncertainties had a highly visible impact on the SMEFT analysis [75]. In particular, they lead to a marked shift in the most likely values of Wilson coefficients while leaving the width of the limits unchanged.

Here, we assess the impact of correlating systematic uncertainties in the top sector. In Fig. 4.11, we show two sets of constraints for a selection of Wilson coefficients. In blue, we show the constraints from a global analysis where all correlations between experimental systematics and theory uncertainties are included. In black, we show the same results, but we treat all experimental systematics as uncorrelated. For all Wilson coefficients, we find good agreement, which indicates that in the top sector, the correlations of systematic uncertainties cannot be ignored but have a limited effect on the final SMEFT limits.

We know that statistical uncertainties are not the leading challenge for global SMEFT analyses. So, if the correlations between experimental systematics are not really relevant either, it leaves the open question of which uncertainties then actually dominate the SMEFT analysis. While for the Higgs sector, the modeling of theory uncertainties has surprisingly little effect on the SMEFT limits [75], the QCD nature of top pair production suggests that the situation will be different here. As a test, in Fig. 4.12, we repeat the comparison of Fig. 4.11, neglecting all theory uncertainties. As before, we show the global analysis with correlated systematics in blue, while in black, these correlations are removed. Now, we are able to see a significant difference. When neglecting the correlations, we observe an increase in the size of the constraints as well as a sizable shift in the most likely point due to the flat distribution modeling theory uncertainties. This is particularly marked in the 2-dimensional constraints on C_{Qq}^{18} and C_{tq}^{8}.

Comparing Figs. 4.11 and 4.12, we learn of the importance of theory uncertainties in the top sector. If we neglect the theory uncertainties, the effect of correlated systematics is non-negligible. While theory uncertainties currently wash out these effects, we expect

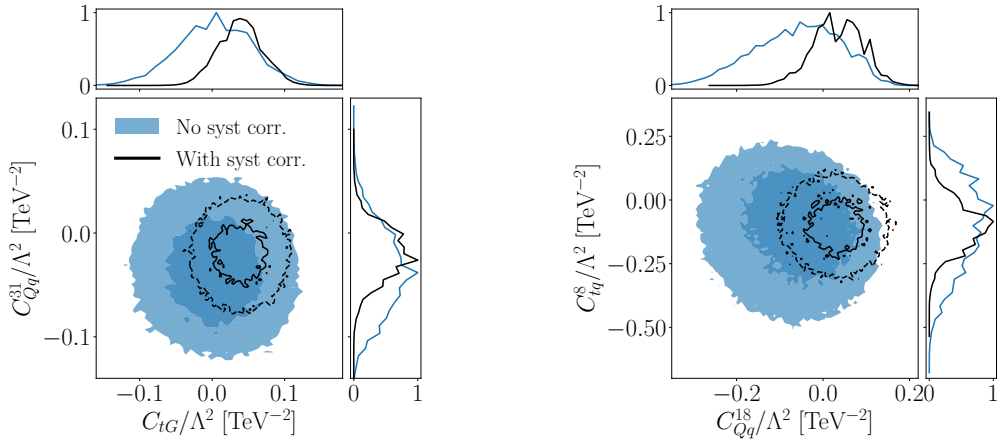

Figure 4.12: Profile likelihood correlations, ignoring theory uncertainties altogether, and either including correlated systematic uncertainties (blue) or ignoring correlations between experimental systematics (black).

them to become more important as SM calculations become more precise and theory uncertainties are reduced. Moreover, we cannot make any statement about the potential impact of public likelihoods for those kinematic measurements that drive the SMEFT sensitivity.

4.3.3 Marginalization vs profiling

As defined in Eq.(4.2), the central object of any SFitter analysis is a fully exclusive likelihood, which is evaluated over the combined space of Wilson coefficients and nuisance parameters. Obviously, the nuisance parameters are irrelevant to the physics interpretation of the global SMEFT analysis. In addition, we are usually not interested in showing all 22 Wilson coefficients at the same time, and instead, we usually reduce this space down to one or two dimensions. Statistically, this can be done by profiling or marginalizing the likelihood. Only for perfect Gaussian distributions do the two methods give the same results, as discussed in Sec. 4.1.2 and Ch. 3. In the Higgs-electroweak sector, significant deviations between the two methods appear through a large under-fluctuation in one bin of a kinematic distribution [75].

Fig. 4.13 displays a selection of correlations from a marginalization (black) and profiling (blue) of the fully exclusive likelihood from all top sector measurements and Wilson coefficients. We show constraints for $\mathcal{O}_{tG}$ vs $\mathcal{O}^1_{tq}$ and $\mathcal{O}^{18}_{Qq}$ vs $\mathcal{O}^8_{Qd}$, but similar effects can be seen in many operator pairs. In general, marginalization leads to narrower constraints than results obtained from profiling. This is particularly evident in the left panel of Fig. 4.13, and it is due to theory uncertainties and their flat likelihood distribution. With this choice, the profile likelihood can force a perfect agreement between data and predictions over a wide range of values for critical Wilson coefficients. When we marginalize over the exclusive likelihood, the difference between Gaussian and flat uncertainties is less pronounced, leading to more Gaussian and narrower one-dimensional distributions, as discussed in detail in Ref. [75]. This effect is especially visible in the top sector, where theory uncertainties are not only poorly defined [223], but also large.

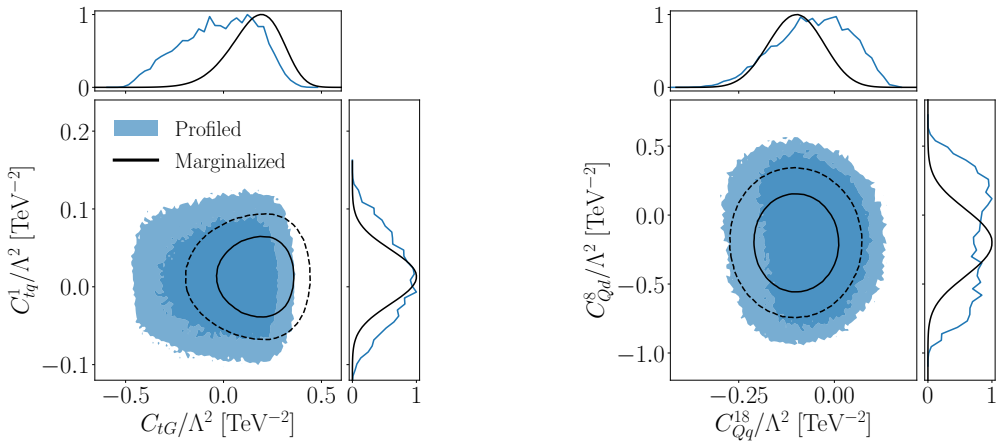

Figure 4.13: Comparison between marginalization (black) and profiling (blue) in a global analysis of the top sector.

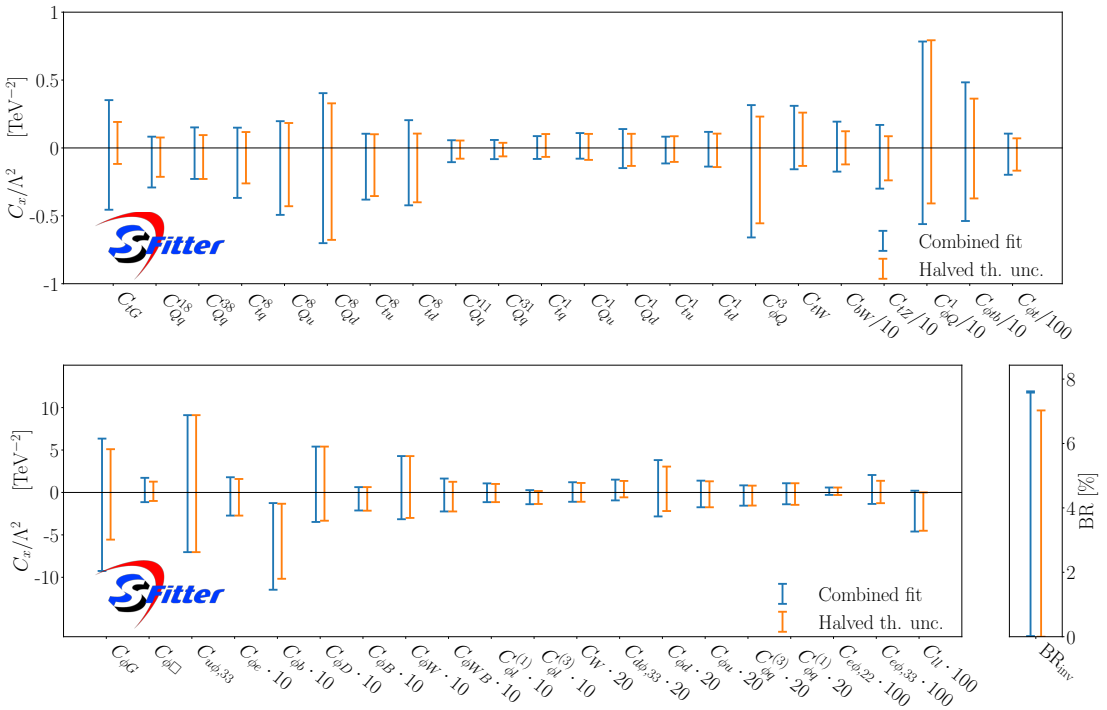

Figure 4.14: Results from a combined SMEFT analysis of the top sector and the Higgs-electroweak sector, showing the constraints at 95% CL on 43 degrees of freedom, resulting from a profiled likelihood.

4.3.4 Top-Higgs-electroweak combination

Finally, making use of the numerical improvements in the SFitter implementation, we can combine the top-sector SMEFT analysis from this paper with the SFitter analysis of the Higgs, di-boson, and electroweak precision observables from Ref. [75]. This combination has been studied in the literature in detail, showing that the two sectors are linked, for instance, through $\mathcal{O}_{tG}$ [138, 139].

We confirm this state of the art and show the combined SFitter profile likelihood of the two sectors in Fig. 4.14. In total, 43 degrees of freedom are constrained: The 22 coefficients constrained by the top sector and 21 additional operators relevant to the Higgs, di-boson, and electroweak observables, discussed in Sec 2.1.2, where the notation and conventions for both operator sets are provided. From the discussion above and in Ref. [75], it is clear that the challenges and limitations of the global analyses in the two sectors are not the same. We show the limits at 95% CL from one-dimensional profile likelihoods of the combined fit (blue) and under the assumption of theory uncertainties reduced by a factor of 2 (orange). The numerical values of the constraints shown in Fig. 4.14 are provided in Tab. A.1 in App. A.

In the top sector, we find strong constraints on the four-fermion operators. The constraints on their Wilson coefficients are driven by kinematic distributions such as the ATLAS measurement of the boosted top discussed in Sec. 4.1.1, and therefore, theory uncertainties do not play an important role in their constraints. Conversely, the constraint on C_{tG} improves significantly when theory uncertainties are halved, indicating that theory uncertainties dominate constraints obtained from top quark pair production total cross sections. Similarly, this hypothetical reduction of theory uncertainties has an effect on

the constraints obtained from the single top, $t\bar{t}W$ and $t\bar{t}Z$ on coefficients such C_{tW}, C_{bW}, and C_{tZ}.

On the other hand, we observe no significant changes in the constraints on the operators relevant to the Higgs, di-boson, and electroweak sectors, shown in the lower half of Fig. 4.14, when theory uncertainties are reduced. The exception is $C_{\phi G}$, which also benefits from the top quark data through its correlation with C_{tG} and $C_{t\phi}$. This is in agreement with Ref. [75], where it was found that in the Higgs-gauge sector, systematic uncertainties are the dominant source of uncertainty for many of the observables in this sector.

4.4 Outlook

Global SMEFT analyses are an exciting development at the LHC, as they combine their role as a precision hadron collider with the goal of interpreting all measurements in terms of precision quantum field theory. This precision theme implies that even if we know that the current measurements do not rule out the Standard Model, limits on SMEFT Wilson coefficients provide important information about fundamental physics.

To extract limits on fundamental physics parameters, we need a comprehensive uncertainty treatment covering experimental statistical uncertainties, experimental systematics, and theory uncertainties. For the latter two, it is crucial to include correlations. Public likelihoods are the state of the art in communicating such experimental results to a broader community. Therefore, we include, for the first time, public ATLAS likelihoods for cross section measurements in a global analysis. These public likelihoods allow us to systematically evaluate the effects of correlations of systematic and theory uncertainties on a global analysis.

The basis of the global SFitter analysis is a fully exclusive likelihood. It includes a large set of rate and kinematic measurements, either pre-processed by ATLAS or CMS, unfolded, or extracted and backward-engineered from experimental publications. The uncertainty treatment is very flexible, including a choice of flat nuisance parameters for correlated theory uncertainties. Starting from the fully exclusive likelihood, we can employ a profile likelihood or a Bayesian marginalization treatment to extract limits on individual Wilson coefficients. In the top sector, we find no significant difference between these two statistical approaches.

The focus of this chapter was on the role of different uncertainties, their correlations, and the role of public likelihoods in this context. In a similar analysis, albeit without public likelihoods, we found that in the electroweak sector, the correlations were crucial, whereas the theory uncertainties were not (yet) a limiting factor [75]. Intriguingly, the situation in the top sector is the opposite: Theory uncertainties are crucial, while the correlations of experimental systematics have a limited impact on the SMEFT limits. This reflects the QCD nature and the statistics of top pair production.

We have demonstrated that public likelihoods provide a much more flexible approach to handling nuisance parameters. However, fully leveraging their potential currently proves difficult due to the large number of measurements included in our global analysis. We emphasize that this is not a final statement about public likelihoods in SMEFT analyses. The reason for this is that we find kinematic measurements of boosted top pair production driving the improvement of the SMEFT limits. For unfolded kinematic

measurements, there are no public likelihoods available yet, but we are looking forward to implementing them in SFitter in the future once they are available.

We finished this study of the impact of theory uncertainties in a consistent theory framework of LHC data by performing the first combined SFitter analysis of the Higgs, electroweak, and top sectors. This further displayed the limiting effect of theory uncertainties on the constraining power of modern top measurements compared to those in the Higgs sector.

Chapter 5

Global EDM analysis

The research and results presented in this chapter are based on work in collaboration with Skyler Degenkolb, Tanmoy Modak, Margarete Mühlleitner and Tilman Plehn and have been published in Ref. [2]. All figures and tables, as well as parts of the text, are similar or identical to the ones in the article.

We have introduced the concepts of EFTs and how EDMs can be addressed as a low-energy EFT in Ch. 2. Based on this theoretical approach, the next step is to implement the EDM framework into SFitter as a tool for global analyses, discussed in Ch. 3. Based on an MCMC algorithm and equipped with a frequentist approach for likelihood construction, SFitter is ideal for the task of performing a global analysis using EDM data. Therefore, it has to be adapted from SMEFT data and parameters, coming with their own challenges, to also perform an analysis using EDM data. Different from the linear and squared contributions in SMEFT, the relation between EDM measurements and their model parameters is linearly, but the uncertainty implementation differs slightly. For LHC experiments, all uncertainties are treated as uncertainties on the experimental data, whereas now only systematic and statistical uncertainties are applied to experimental data, while the theory uncertainties depend on the model parameters. This global analysis provides updated limits on a multidimensional model space and a comprehensive uncertainty treatment, including uncertainties from theoretical calculations. Also, we can identify the impact of current and future measurements on the parameter limits or the impact of reduced uncertainties due to the flexibility of SFitter.

This chapter is organized in the following way: We start by describing the used measurements, their implementation, uncertainties, and the model parameters entering the considered processes in Sec. 5.1. Next, we provide single parameter ranges from the data on individual model parameters in Sec. 5.2. Then, Sec. 5.3 discusses the results from the global analysis with and without theory uncertainties, first, on a well-defined sub-space and then on the full parameter configuration.

5.1 EDM Measurements

The Lagrangian parameters derived in Section 2.2, Eq. (2.42), can be used to predict the measured EDMs d_i. These measurements are linked to the Lagrangian parameters in terms of linear combinations with system-specific coefficients α_{i,c_j},

$$d_i = \sum_{c_j} \alpha_{i,c_j} c_j \; . \tag{5.1}$$

System i	Measured d_i [e cm]	Upper limit on $\lvert d_i\rvert$ [e cm]	Reference
n	$(0.0 \pm 1.1_{\text{stat}} \pm 0.2_{\text{syst}}) \cdot 10^{-26}$	$2.2 \cdot 10^{-26}$	[230]
^{205}Tl	$(-4.0 \pm 4.3) \cdot 10^{-25}$	$1.1 \cdot 10^{-24}$	[231]
^{133}Cs	$(-1.8 \pm 6.7_{\text{stat}} \pm 1.8_{\text{syst}}) \cdot 10^{-24}$	$1.4 \cdot 10^{-23}$	[232]
HfF^+	$(-1.3 \pm 2.0_{\text{stat}} \pm 0.6_{\text{syst}}) \cdot 10^{-30}$	$4.8 \cdot 10^{-30}$	[233]
ThO	$(4.3 \pm 3.1_{\text{stat}} \pm 2.6_{\text{syst}}) \cdot 10^{-30}$	$1.1 \cdot 10^{-29}$	[234]
YbF	$(-2.4 \pm 5.7_{\text{stat}} \pm 1.5_{\text{syst}}) \cdot 10^{-28}$	$1.2 \cdot 10^{-27}$	[235]
^{199}Hg	$(2.20 \pm 2.75_{\text{stat}} \pm 1.48_{\text{syst}}) \cdot 10^{-30}$	$7.4 \cdot 10^{-30}$	[236, 237]
^{129}Xe	$(-1.76 \pm 1.82) \cdot 10^{-28}$	[238, 239]	
^{171}Yb	$(-6.8 \pm 5.1_{\text{stat}} \pm 1.2_{\text{syst}}) \cdot 10^{-27}$	$1.5 \cdot 10^{-26}$	[240]
^{225}Ra	$(4 \pm 6_{\text{stat}} \pm 0.2_{\text{syst}}) \cdot 10^{-24}$	$1.4 \cdot 10^{-23}$	[241]
TlF	$(-1.7 \pm 2.9) \cdot 10^{-23}$	$6.5 \cdot 10^{-23}$	[242]
	Measured ω_i [mrad/s]	Rescaling factor x_i for d_i	Reference
HfF^+	$(-0.0459 \pm 0.0716_{\text{stat}} \pm 0.0217_{\text{syst}})^\dagger$	0.999	[233]
ThO	$(-0.510 \pm 0.373_{\text{stat}} \pm 0.310_{\text{syst}})$	0.982	[234]
YbF	$(5.30 \pm 12.60_{\text{stat}} \pm 3.30_{\text{syst}})$	1.12	[235]

Table 5.1: Measured EDM values and 95%CL ranges. For ^{129}Xe we combine two independent results with similar precision, using inverse-variance weighting. For the open-shell molecules, we also provide the measured angular frequencies and the rescaling factor which allows us to use $x_i d_i$ for each experimentally reported d_i. For the definition of x_i, see text. †The frequency for HfF^+ is scaled by a factor of 2 relative to Ref. [233], to consistently use Eq. (5.8) for all systems.

The measurements we analyze are listed in Tab. 5.1 and discussed below. Unless otherwise indicated, we always refer to the isotopes and charge states that are given in Tab. 5.1, with the relevant isotopes, for the considered molecular systems in this global analysis, these are ^{180}Hf, ^{232}Th, ^{174}Yb, ^{205}Tl, ^{16}O, and ^{19}F. Furthermore, we decided to neglect constraints of weaker experimental bounds from ^{85}Rb [243, 243], Xe^m [244], PbO [245], $Eu_{0.5}Ba_{0.5}TiO_3$ [246], and the Λ hyperon [247]. Additionally, the limits from μ and τ leptons will factorize from the hadronic-scale Lagrangian and thus will not be included in this global analysis, but will be part of future analyses.

As discussed in Sec. 2.2, the α-values for all $C_{S,P,T}^{(0,1)}$ can be extracted from the corresponding relations of the low-energy parameters given in Eq. (2.42). For this derivation we follow the steps from Eqs. (2.35)-(2.38) and find

$$\begin{aligned}
\alpha_{C_S^{(0)}} &= \alpha_{C_S} - \alpha_{C_P} \frac{g_P^{(1)}}{g_S^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle} \\
\alpha_{C_P^{(0)}} &= \alpha_{C_P} + \alpha_{C_S} \frac{g_S^{(1)}}{g_P^{(0)}} \frac{Z-N}{Z+N} \\
\alpha_{C_T^{(0)}} &= \left(1 - \frac{g_T^{(1)}}{g_T^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle}\right) \alpha_{C_T} \,,
\end{aligned} \tag{5.2}$$

System i	$\langle\sigma_n\rangle$	$\langle\sigma_p\rangle$	$\langle\sigma_z\rangle^{(0)}$	α_{i,C_S} [e cm]	α_{i,C_P} [e cm]	α_{i,C_T} [e cm]
Tl	0.274	0.726	1	$-6.77 \cdot 10^{-18}$ [249]	$1.5 \cdot 10^{-23}$ [250]	$5 \cdot 10^{-21}$ [250]
Cs	−0.206	−0.572	−0.778	$7.8 \cdot 10^{-19}$ [24]	$2.2 \cdot 10^{-23}$ [250]	$9.2 \cdot 10^{-21}$ [250]
^{199}Hg	−0.302	−0.032	−0.334	$-2.8 \cdot 10^{-22}$ [251]	$6 \cdot 10^{-23}$ [252]	$1.7 \cdot 10^{-20}$ [252]
^{129}Xe	0.73	0.27	1	$-6.28 \cdot 10^{-23}$ [252]	$1.6 \cdot 10^{-23}$ [252]	$5.7 \cdot 10^{-21}$ [252]
^{171}Yb	−0.3	−0.034	−0.334	$-7.34 \cdot 10^{-22}$ [253]	$3.60 \cdot 10^{-23}$ [253]	$1.04 \cdot 10^{-20}$ [253]
^{225}Ra	0.72	0.28	1	$5.63 \cdot 10^{-21}$ [253]	$-6.4 \cdot 10^{-22}$ [252]	$-1.8 \cdot 10^{-19}$ [252]
TlF	0.274	0.726	1	$1.09 \cdot 10^{-16}$ [253]	$3.8 \cdot 10^{-18}$ [253]	$1.06 \cdot 10^{-15}$ [254]

Table 5.2: Effective parameters used as input to our global analysis, as summarized in Tab. 5.3.

with the $\alpha_{C_{S,P,T}}$ given in Tab. 5.2. In this table we also show

$$\langle\sigma_z\rangle^{(0)} = \langle\sigma_n\rangle + \langle\sigma_p\rangle \ , \tag{5.3}$$

which is proportional to the isoscalar sum of neutron and proton spin projections in a shell model of the nucleus, a factor of two being given by the usual relation of the spin and Pauli operators. The shell model is not expected to be reliable for the deformed nuclei ^{171}Yb and ^{225}Ra. The values given in Tab. 5.2 are chosen to optimize the agreement of the calculated and measured nuclear magnetic moments within this framework. Literature values are available for these nuclear species [91]. The spin fractions contributing to semileptonic coefficients in TlF only take into account the ^{205}Tl nucleus, though see Ref. [248] for some consideration of contributions from the ^{19}F nucleus. For $\alpha_{^{129}\mathrm{Xe},C_S}$ we use a scaling relation to derive a value from $\alpha_{^{129}\mathrm{Xe},C_T}$ of the cited reference.

5.1.1 Nucleons

Starting with the simplest possible, hadronic EDM measurement, from single nucleons. Their EDM can be translated directly from the Lagrangian in Eq. (2.28) and written as

$$-\frac{i}{2}F^{\mu\nu}d_N\ \left(\bar{N}\sigma_{\mu\nu}\gamma_5 N\right) \ . \tag{5.4}$$

Despite the direct translation from the Lagrangian, the nucleon EDM can also be described by chiral perturbation theory, which is based on the hadronic-scale Lagrangian. In that case, the nucleon EDM consists of short-range nucleon contributions d_N^{sr}, NNLO pion-loop contributions, and potential direct contributions to Eq. (5.4) within and beyond SM [255],

$$\begin{aligned} d_n &= d_n^{\mathrm{sr}} - \frac{eg_A}{8\pi^2 F_\pi}\left[g_\pi^{(0)}\left(\ln\frac{m_\pi^2}{m_N^2} - \frac{\pi m_\pi}{2m_N}\right) - \frac{g_\pi^{(1)}}{4}\left(\kappa_0-\kappa_1\right)\frac{m_\pi^2}{m_N^2}\ln\frac{m_\pi^2}{m_N^2}\right] \\ d_p &= d_p^{\mathrm{sr}} + \frac{eg_A}{8\pi^2 F_\pi}\left[g_\pi^{(0)}\left(\ln\frac{m_\pi^2}{m_N^2} - \frac{2\pi m_\pi}{m_N}\right) - \frac{g_\pi^{(1)}}{4}\left(\frac{2\pi m_\pi}{m_N} + \left(\frac{5}{2}+\kappa_0+\kappa_1\right)\frac{m_\pi^2}{m_N^2}\ln\frac{m_\pi^2}{m_N^2}\right)\right], \end{aligned} \tag{5.5}$$

where $F_\pi = 92$ MeV is the pion decay constant [256], $m_\pi = 139$ MeV, $m_N = 940$ MeV, $g_A \approx 1.27$ is the nucleon isovector axial charge, and the isoscalar and isovector nucleon anomalous magnetic moments are $\kappa_0 = -0.12$ and $\kappa_1 = 3.7$.

The renormalization scale in this case is set to the nucleon mass, while the occurrence of splitting the nucleon into proton and neutron masses will lead to higher-order effects. However, these effects and uncertainties coming from higher-order contributions are negligibly small compared to other sources of experimental and theoretical uncertainties. The hadronic-scale parameters can be related to parameters of the weak-scale, where the finite values of $g_\pi^{(0,1)}$ are related to a CKM phase or $\bar{\theta}$, but also to hadronic 4-fermion operators. The input from the presently most sensitive neutron EDM measurement to our global analysis is given in Tab. 5.1.

In Eq. (5.5) we can see, that we could either use $d_{n,p}^{\text{sr}}$ or $d_{n,p}$ in our set of model parameters. In terms of a global analysis, the choice does not matter, because these Lagrangian parameters, including their renormalization condition, are just the parameters of the outcome. Nevertheless, their real purpose is to compute other observables and compare them to measurements using a consistent theoretical framework. Choosing one definition will thus have an impact on the definition of other parameters.

As argued in Sec. 2.2, one could absorb d_p or equivalently d_p^{sr} into d_n, d_n^{sr} respectively. Then, the two choices will no longer be equivalent, and we must be careful when switching from one version to another. Choosing $d_{n,p}$ as model parameters implies that we use measured values as Lagrangian parameters, with a corresponding renormalization condition called on-shell renormalization in collider physics. In that case, we extract d_n and d_p from data and could use Eq. (5.5) to translate them into $d_{n,p}^{\text{sr}}$. The $d_{n,p}$ scheme is used throughout this thesis, however the alternative implementation using $d_{n,p}^{\text{sr}}$, which is a better approximation for the relation $d_p^{\text{sr}} = -d_n^{\text{sr}}$, will be discussed in App. B, where we also present a complete set of results using this approach.

5.1.2 Open-shell (paramagnetic) systems

Paramagnetic atoms and molecules, more precisely referred to as atoms and molecules with open electronic shells[1], are primarily sensitive to the electron EDM and scalar electron-nucleon couplings $C_S^{(0,1)}$, written down in Eq. (2.30), due to their atomic shell structure. For open-shell systems, it is common to distinguish between the case of open-shell atoms with the atomic EDM contribution

$$d_i = \alpha_{i,d_e} d_e + \alpha_{i,C_S} C_S + \sum_{c_j} \alpha_{i,c_j} c_j \qquad \text{for} \qquad i \in \{\text{Tl}, \text{Cs}\} \tag{5.6}$$

and open-shell molecules. Nevertheless, in both cases, the terms involving the Lagrangian parameters d_e and $C_S^{(0,1)}$ dominate. For molecular systems, the experimentally relevant or measured quantity is not a fixed dipole moment but rather a phase difference that accumulates over the measurement time and is interpreted as frequency. This frequency is often reported as a parity- and time-reversal-violating frequency shift ω_i,

$$\omega_i = \eta_{i,d_e}^{(m)} d_e + k_{i,C_S}^{(m)} C_S + \sum_{c_j} \alpha_{i,c_j}^{(m)} c_j \qquad \text{for} \qquad i \in \{\text{HfF}^+, \text{ThO}, \text{YbF}\}\,, \tag{5.7}$$

where the superscript (m) indicates the molecular system. These angular frequencies are also listed in Tab. 5.1. Given ω_i as frequency, emphasize that it does not scale

[1] We prefer the nomenclature of *open-* and *closed-shell* systems, rather than paramagnetic or diamagnetic (respectively), since this more clearly indicates the properties that are relevant for determining the leading contributions to the total system EDM.

System i	α_{i,d_e}	$\alpha_{i,C_S^{(0)}}$ [e cm]	$\alpha_{i,C_P^{(0)}}$ [e cm]	$\alpha_{i,C_T^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(1)}}$ [e cm]	α_{i,d_n}	α_{i,d_p}
n	–	–	–	–	0	0	1	(-1)
^{205}Tl	$-558^{\pm 28}$ [249]	$-6.77^{\pm 0.34} \cdot 10^{-18}$	$1.4^{+2.5}_{-0.8} \cdot 10^{-19}$	$8.8^{+4.0}_{-1.2} \cdot 10^{-21}$	$-6.41^{+3.57}_{-4.51} \cdot 10^{-18}$	$2.27^{+6.13}_{-1.19} \cdot 10^{-19}$	$-5.75^{+3.44}_{-4.44} \cdot 10^{-6}$	$1.61^{+3.35}_{-7.56} \cdot 10^{-5}$
^{133}Cs	$123^{\pm 4}$	$7.80^{+0.2}_{-0.8} \cdot 10^{-19}$	$-1.4^{+0.8}_{-2.2} \cdot 10^{-20}$	$1.7^{\pm 0.4} \cdot 10^{-20}$	$8.09^{+151.1}_{-2.63} \cdot 10^{-19}$	$2.70^{+3.27}_{-1.79} \cdot 10^{-18}$	$1.0^{\pm 1} \cdot 10^{-5}$	$6.7^{\pm 2} \cdot 10^{-4}$
^{199}Hg	$11.6^{+10}_{-18} \cdot 10^{-3}$ [253, 257]	$-1.26^{+0.7}_{-1.6} \cdot 10^{-21}$	$6.6^{+2.4}_{-1.6} \cdot 10^{-23}$	$-6.4^{+3}_{-4} \cdot 10^{-21}$	$-3.05^{+1.75}_{-14.5} \cdot 10^{-18}$	$-6.10^{+16.6}_{-25.5} \cdot 10^{-18}$	$-1.36^{+0.38}_{-3.45} \cdot 10^{-4}$	$-1.36^{+0.38}_{-5.12} \cdot 10^{-5}$
^{129}Xe	$-8^{+0}_{-8} \cdot 10^{-4}$ [253, 257]	$-2.1^{+1.2}_{-2.5} \cdot 10^{-22}$	$1.7^{+0.5}_{-0.4} \cdot 10^{-23}$	$1.24^{+0.78}_{-0.61} \cdot 10^{-21}$	$-3.91^{+1.75}_{-23.4} \cdot 10^{-19}$	$2.93^{+24.4}_{-1.64} \cdot 10^{-19}$	$2.29^{+0.77}_{-0.58} \cdot 10^{-5}$	$4.89^{+1.64}_{-1.25} \cdot 10^{-6}$
^{171}Yb	$1.44^{+165}_{-4.5} \cdot 10^{-3}$ [253]	$-1.31^{+0.55}_{-1.06} \cdot 10^{-21}$	$4.93^{+3.54}_{-1.55} \cdot 10^{-23}$	$-3.68^{+1.86}_{-2.43} \cdot 10^{-21}$	$-2.84^{+0.43}_{-6.05} \cdot 10^{-18}$	$-5.67^{+7.74}_{-10.3} \cdot 10^{-18}$	$-1.13^{+0.32}_{-0.28} \cdot 10^{-4}$	$-1.13^{+0.32}_{-0.28} \cdot 10^{-5}$
^{225}Ra	$-5.4^{+134}_{-2.0} \cdot 10^{-2}$ [253]	$1.13^{+2.94}_{-0.51} \cdot 10^{-20}$	$-7.63^{+2.05}_{-3.88} \cdot 10^{-22}$	$-4.5^{+2.0}_{-2.5} \cdot 10^{-20}$	$1.72^{+5.73}_{-0.67} \cdot 10^{-15}$	$-6.89^{+2.66}_{-22.9} \cdot 10^{-15}$	$-5.38^{+1.20}_{-1.58} \cdot 10^{-4}$	$-1.19^{+0.27}_{-0.35} \cdot 10^{-4}$
TlF	$1.36^{+0.36}_{-0.32} \cdot 10^{3}$ [253]	$1.44^{+0.8}_{-0.5} \cdot 10^{-16}$	$1.51^{+2.2}_{-5.6} \cdot 10^{-18}$	$1.87^{+0.19}_{-0.17} \cdot 10^{-15}$	$-7.00^{+1.39}_{-2.17} \cdot 10^{-13}$	$2.48^{+4.58}_{-0.34} \cdot 10^{-14}$	$-6.28^{1.71}_{-2.28} \cdot 10^{-1}$	$1.76^{+2.41}_{-6.76}$
HfF^+	1	$9.17^{\pm 0.06} \cdot 10^{-21}$	–	–	–	–	–	–
ThO	1	$1.51^{+0}_{-0.2} \cdot 10^{-20}$	–	–	–	–	–	–
YbF	1	$8.99^{\pm 0.70} \cdot 10^{-21}$	–	–	–	–	–	–
	$\eta_{i,d_e}^{(m)} \left[\frac{\text{mrad}}{\text{s } e \text{ cm}}\right]$	$k_{i,C_S}^{(m)} \left[\frac{\text{mrad}}{\text{s}}\right]$	α_{i,C_P}	α_{i,C_T}	$\alpha_{i,g_\pi^{(0)}}$	$\alpha_{i,g_\pi^{(1)}}$	α_{i,d_n}	α_{i,d_p}
HfF^+	$3.49^{\pm 0.14} \cdot 10^{28}$ [251, 258–261]	$3.2^{+0.1}_{-0.2} \cdot 10^{8}$ [251, 258, 259]	–	–	–	–	–	–
ThO	$-1.21^{+0.05}_{-0.39} \cdot 10^{29}$ [251, 262–264]†	$-1.82^{+0.42}_{-0.27} \cdot 10^{9}$ [251, 262, 264–266]†	–	–	–	–	–	–
YbF	$-1.96^{\pm 0.15} \cdot 10^{28}$ [251, 265–268]	$-1.76^{\pm 0.2} \cdot 10^{8}$ [251, 265–267]	–	–	–	–	–	–

Table 5.3: Central values and theory uncertainties for the α-parameters defined in Eq. (5.1), using $d_{n,p}$ as model parameters. The parentheses around α_{n,d_p} indicate our assumption Eq. (2.41), i.e., this value should not be construed as entering Eq. (5.1). (For the implementation using $d_{n,p}^{\text{sr}}$, see Appendix B and in particular Tab.B.2.) Here, a "$-$" means that we neglect the dependence in our global analysis. †There appears to be an overall sign issue in the coefficients reported for ThO in Table 4 of Ref. [251]. The values for Tl and Cs $\alpha_{i,C_T^{(0)}}$ are estimated by simple analytical calculations [250], and the uncertainties quoted here are estimated as approximately twice those arising from the relevant hadronic matrix elements.

with the experimental applied electric field in a simple way, as it is the case for atomic structures. It rather depends on the specific molecular structure and the so-called effective electric field E_{eff} that saturates when the molecule is polarized. It is noticeably, that α_{i,d_e} is dimensionless, while $\eta^{(m)}_{i,d_e}$ has dimension $[\text{mrad} \times (\text{s } e \text{ cm})^{-1}]$. Similarly, α_{i,C_S} has the units of an EDM with $[e \text{ cm}]$, while $k^{(m)}_{i,C_S}$ has units of angular frequency given by $[\text{mrad} \times (\text{s})^{-1}]$.

Considering the remaining contributions to open-shell molecular EDMs, which have weak dependencies on other low-energy constants, we can see that, in particular, the μ and τ leptons are indirectly constrained via the limits from ThO and Hg [269]. However, we are not aware of any established values for the coefficients of semileptonic or hadronic parameters. We thus take $\alpha^{(m)}_{i,c_j} = 0$ in Eq. (5.7), while for the open-shell atoms we obtain values for semileptonic coefficients either from the literature or by scaling arguments, as reported in Tabs. 5.2 and 5.3.

Starting from the truncated version of Eq. (5.7), we are able to adapt the convention in such a way that all signs, g-factors, spin magnitudes etc. are absorbed into the coefficients on the right side, leaving us with

$$\begin{aligned} \omega_i &= \eta^{(m)}_{i,d_e} d_e + k^{(m)}_{i,C_S} C_S \\ &= -\frac{E_{\text{eff},i}}{\hbar} d_e + \frac{W_i}{\hbar} C_S \,. \end{aligned} \tag{5.8}$$

Commonly used, $\mathcal{E}_{\text{eff}}$ refers to the effective electric field where

$$E_{\text{eff}} = \mathcal{E}_{\text{eff}} \,\text{sgn}(\vec{J} \cdot \hat{n}) \,\langle \hat{n} \cdot \hat{z} \rangle \,, \tag{5.9}$$

and $\vec{J}$ is the total electronic angular momentum, $\hat{n}$ is the direction of the internuclear axis, and $\hat{z}$ is the direction of the externally applied electric field. With this we can define $\Omega = \vec{J} \cdot \hat{n}$ as a quantum number, and $\langle \hat{n} \cdot \hat{z} \rangle$ indicates the degree of polarization of the molecule by the applied electric field. More information and conventions used for the internal electric field, orientation of the internuclear axis, etc. can be found in Appendix A.3 of Ref. [270].

The truncated expression for the frequency shift thus provides us with physical expressions for the two constants,

$$\eta^{(m)}_{i,d_e} = -\frac{E_{\text{eff},i}}{\hbar} \qquad \text{and} \qquad k^{(m)}_{i,C_S} = \frac{W_i}{\hbar} \,. \tag{5.10}$$

Unfortunately, the frequency measurement cannot be directly converted into a permanent molecular dipole moment due to the molecule's electrical polarization during the measurement. The linear energy shift originates in the saturation of the electric field in the lab-frame given by $\langle \hat{n} \cdot \hat{z} \rangle \to 1$, and the limited dependency on the external field is the same as the one of an induced dipole moment. Therefore, to translate the molecular frequency difference ω_i to terms of a permanent dipole moment, we define the molecule's d_i in relation to the given effective electric field,

$$-E_{\text{eff},i} d_i \equiv -E_{\text{eff},i} d_e + W_i C_S \,. \tag{5.11}$$

We can use this form to define, in analogy to Eq. (5.6),

$$d_i = d_e + \alpha_{i,C_S} C_S \qquad \Leftrightarrow \qquad \alpha_{i,d_e} = 1 \quad \text{and} \quad \alpha_{i,C_S} = -\frac{W_i}{E_{\text{eff},i}} = \frac{k^{(m)}_{i,C_S}}{\eta^{(m)}_{i,d_e}} \,. \tag{5.12}$$

With this treatment, we can also re-cast the limits from the open-shell molecules for our global analysis in units of [e cm]. Now, all coefficients for all considered systems are expressed in the same units, which has the side effect, that $\alpha_{i,d_e} = 1$ holds for all open-shell molecules. This approach also serves two additional purposes: (i) Many different conventions are used for these coefficients, sometimes using the same symbols to describe different quantities. In many publications, these measured quantities are linked to the electron EDM, which makes a comparison relatively straightforward and all quantities have to be ultimately connected to a measured experimental phase; (ii) As first pointed out in Refs. [265, 266], there is a considerable variation in the literature for given values of $\eta^{(m)}_{i,d_e}$ and $k^{(m)}_{i,C_S}$ in a given system, but there is much less variation in their ratio. By dividing Eq. (5.7) by $\eta^{(m)}_{i,d_e}$ the uncertainty on this coefficient should be passed on to all other semileptonic and hadronic coefficients.

Note that we use $\eta^{(m)}_{i,d_e}$ to obtain the sole-source limits on d_e from the experimentally measured ω_i but also to convert the experimentally measured frequency values into EDM units for Tab. 5.1. These experimental limits for EDMs from open-shell molecules in Tab. 5.1 are all rescaled by the indicated factor x_i. The listed x_i differ from the unity value in the case when the updated values for $\eta^{(m)}_{i,d_e}$ differ from the cited publications. This is typically the case when improved molecular structure calculations have become available since the experimental limit was published.

Recommended values for many of the α_{i,c_j} are given in Tables III-V of Ref. [26] and in Table 4 of Ref. [251], including in many cases the ranges corresponding to theory uncertainties (or at least different reported values). We give our choices for α_{i,c_j} in Tab. 5.3.

5.1.3 Closed-shell (diamagnetic) systems

In contrast to the open-shell systems, where sub-leading contributions are largely neglected due to lack of theory inputs, the small contributions of d_e and $C_S^{(0)}$ are considered for all closed-shell systems. These are typically smaller contributions to the EDM observable, mostly in such cases where all electron spins are paired and the main contributions coming rather from nucleon EDMs, nuclear forces, which are mediated by pion exchange with strengths $g_\pi^{(0,1,2)}$, and the nuclear-spin-dependent semileptonic interactions $C_T^{(0)}$ (and possibly $C_P^{(0)}$). The experimental precision, especially of Hg, is nevertheless high enough to contribute meaningful constraining power for d_e and $C_S^{(0)}$.

Unfortunately, it seems, that the arguments [251] towards the effect of closed-shell systems and their constraining power in the $d_e - C_S$ subspace due to a different sign of the ratio $\alpha_{i,d_e}/\alpha_{i,C_S}$ compared to open-shell systems, does not hold for Hg or most of the other systems listed in Tab. 5.1. While recent calculations [253] may indicate that α_{i,d_e} and α_{i,C_S} in some highly-correlated closed-shell systems could indeed have a different sign, the theory uncertainties for these cases are large enough to cross zero. Numerical calculations for specific systems so far only support ratios of the same sign [271]. Two subtleties are important to consider in this context: (1) The value for α_{i,d_e} includes contributions from

multiple different interactions [252, 253, 257, 272] whose contributions must be summed, and (2) our implementation of $\alpha_{C_S^{(0)}}$ could generate sign changes via the difference of terms in Eq. (5.2).

At the moment, as it can be seen from Tabs. 5.2 and 5.3, Ra and Yb are the only measured systems where this situation appears to arise. These may thus represent special cases among closed-shell systems, in that sense that reducing the error of experimental and theory inputs could significantly impact the $d_e - C_S^{(0)}$ subspace that is complementary to the dominating open-shell constraints. The supporting arguments for such claims should, however, be examined in detail.

The contributions of nuclear forces and nucleon EDMs are frequently interpreted via the Schiff moment of a given nucleus [26], which is more easily related to nuclear structure parameters [250]. In terms of our model coefficients, the Schiff moment S_i of system i and a system-specific coefficient $k_{i,S}$ can be expressed in terms of the corresponding EDM via

$$\begin{aligned} k_{i,S} S_i &= \sum_{c_j \in \{d_{n,p}, g_\pi^{(0,1,2)}\}} \alpha_{i,c_j} c_j \\ &\approx k_{i,S} \left[s_{i,n} d_n + s_{i,p} d_p + \frac{m_N g_A}{F_\pi} \left(a_{i,0} g_\pi^{(0)} + a_{i,1} g_\pi^{(1)} \right) \right] . \end{aligned} \tag{5.13}$$

where the coefficients $s_{i,N}$ $(N = n, p)$ indicate contributions from EDMs of unpaired nucleons and the coefficients $a_{i,m}$ $(m = 0, 1, 2)$ parameterize the strength of a CP-violating pion exchange, organized by isospin, for the nucleus of system i. We drop the $g_\pi^{(2)}$ term as discussed in Chapter 2.2 and note that the coefficients in front of $g_\pi^{(0)}$ and $g_\pi^{(1)}$ will change when we replace $d_{n,p}$ with $d_{n,p}^{\mathrm{sr}}$.

The coefficient $k_{i,S}$ is calculated for many systems of interest and can, in principle, be large in heavy and especially deformed nuclei such as ^{225}Ra. This corresponds to a nuclear-structure induced enhancement of the observable EDM, which applies to both the nucleon EDMs and the pion-exchange forces. This implies that among the potentially leading contributions from the six hadronic-scale model parameters $C_{P,T}^{(0)}$, $d_{n,p}$, and $g_\pi^{(0,1)}$, very different weights can be expected according to the electronic and nuclear structure of the various closed-shell systems. Inspecting Tab. 5.3 reveals that to a limited extent, this is indeed already the case. However, this complementarity is not yet fully exploited.

Despite the pion pole enhancement, $C_P^{(0)}$ appears to be suppressed relative to $C_T^{(0)}$ in all measured systems. The coefficients of $g_\pi^{(0,1)}$ are typically of comparable size for a given system, although possibly different in sign. Unfortunately, the coefficients of $d_{n,p}$ are not well known for most nuclei, although in principle these can be calculated. For the pion-exchange forces, even in nuclei that have been the object of many studies, the corresponding theory uncertainties are large. This is especially true for soft nuclei such as the chosen Hg and Xe, in which non-static deformations can present special challenges.

For nuclei with spin $I > 1/2$, a CP-violating nuclear magnetic Quadrupole moment can, in principle, exist in analogy to a Schiff-moment-induced EDM and can also be analyzed within a common global analysis. At the moment, among the experimental systems that have already been measured, Cs is the only one where this effect is expected to be most relevant.

As part of a global analysis, a large number of experimental measurements from complementary closed-shell systems can disentangle contributions from model parameters, which are relevant at some level for all considered measurements. In this sense the role of d_e and $C_S^{(0)}$ takes on a new importance, not only to further constrain these parameters themselves, but also as an additional contribution to the EDM that brings along uncertainties which dilute the constraining power for other model parameters. Closed-shell molecular systems such as TlF or systems containing Schiff-enhanced nuclides introduce complementary constraining power to our global analysis.

We finally note that closed-shell systems unite theory inputs from different communities. Different conventions for assigning a negative isospin projection in the nucleon doublet must be carefully noted when combining calculated coefficients from different sources, especially when we include semileptonic interactions with nuclear forces. Reference [273] makes an effort to disambiguate a part of this issue for $\alpha_{g_\pi^{(0,1,2)},^{129}\mathrm{Xe}}$.

5.2 Single parameter ranges

In this chapter, we provide toy ranges of individual model parameters taken from the set of

$$\left\{ d_e, C_S^{(0)}, C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n, d_p \right\}, \tag{5.14}$$

using the set of EDM measurements one by one. To extract these single parameter ranges, we utilize Eq. (5.6) and include all EDM measurements listed in Tab. 5.1 and the model dependencies from Tab. 5.3. For the likelihood we assume a Gaussian form, which means that all limits are quoted as symmetric one-sigma error bars around the central, best-fit value. While these single-parameter limits allow us to compare the reach of different measurements for CP violation as a whole, they do not give the allowed ranges of the individual model parameters. Contributions from different model parameters to the same measurement can cancel. Because an EFT builds on the assumption that many higher-dimensional operators are induced by a given physics model, the single-parameter limits are overly optimistic.

Using the single-parameter constraints given in Tab. 5.4 we can compare the impact of different EDM measurements on a given model parameter, with the caveat that multi-dimensional correlations might change that picture. Starting with the electron EDM d_e, the open-shell molecules HfF^+ and ThO provide the strongest constraints, while the open-shell YbF molecule has a similar constraining power as the closed-shell Hg. The same two open-shell molecules lead the constraining power for the scalar electron-nucleon coupling $C_S^{(0)}$, again followed by Hg, YbF, and Tl.

The pseudoscalar and tensor electron-nucleon couplings $C_{P,T}^{(0)}$ can be probed by the open-shell atoms and the closed-shell systems. This is done most efficiently using Hg at present and then, with much reduced constraining power, followed by Tl, Xe, and TlF. It is difficult to estimate the impact of these measurements on the combination of $C_P^{(0)}$ and $C_T^{(0)}$, and we will see in the next section how this more complex dependence affects the full global analysis.

The four hadronic parameters, $g_\pi^{(0,1)}$ and $d_{n,p}$, are most strongly constrained by the neutron EDM measurement and again Hg, suggesting that there will be significant

System i	d_e [e cm]	$C_S^{(0)}$	$C_P^{(0)}$	$C_T^{(0)}$
Tl	$(7.2 \pm 7.7) \cdot 10^{-28}$	$(5.9 \pm 6.4) \cdot 10^{-8}$	$(-2.9 \pm 3.1) \cdot 10^{-6}$	$(-4.5 \pm 4.9) \cdot 10^{-5}$
Cs	$(-1.5 \pm 5.6) \cdot 10^{-26}$	$(-2.3 \pm 8.9) \cdot 10^{-6}$	$(1.3 \pm 5.0) \cdot 10^{-4}$	$(-1.1 \pm 4.1) \cdot 10^{-4}$
^{199}Hg	$(1.9 \pm 2.7) \cdot 10^{-28}$	$(-1.7 \pm 2.5) \cdot 10^{-9}$	$(3.3 \pm 4.7) \cdot 10^{-8}$	$(-3.4 \pm 4.9) \cdot 10^{-10}$
^{129}Xe	$(2.2 \pm 2.3) \cdot 10^{-25}$	$(8.4 \pm 8.7) \cdot 10^{-7}$	$(-1.0 \pm 1.1) \cdot 10^{-5}$	$(-1.4 \pm 1.5) \cdot 10^{-7}$
^{171}Yb	$(-4.7 \pm 3.6) \cdot 10^{-24}$	$(5.2 \pm 4.0) \cdot 10^{-6}$	$(-1.4 \pm 1.1) \cdot 10^{-4}$	$(1.8 \pm 1.4) \cdot 10^{-6}$
^{225}Ra	$(-0.7 \pm 1.1) \cdot 10^{-22}$	$(3.5 \pm 5.3) \cdot 10^{-4}$	$(-5.2 \pm 7.9) \cdot 10^{-3}$	$(-0.9 \pm 1.3) \cdot 10^{-4}$
TlF	$(-1.3 \pm 2.1) \cdot 10^{-26}$	$(-1.2 \pm 2.0) \cdot 10^{-7}$	$(-1.1 \pm 1.9) \cdot 10^{-5}$	$(-0.9 \pm 1.6) \cdot 10^{-8}$
HfF^+	$(-1.3 \pm 2.1) \cdot 10^{-30}$	$(-1.4 \pm 2.3) \cdot 10^{-10}$		
ThO	$(4.3 \pm 4.0) \cdot 10^{-30}$	$(2.8 \pm 2.7) \cdot 10^{-10}$		
YbF	$(-2.4 \pm 5.9) \cdot 10^{-28}$	$(-2.7 \pm 6.6) \cdot 10^{-8}$		

	$g_\pi^{(0)}$	$g_\pi^{(1)}$	d_n [e cm]	d_p [e cm]
Tl	$(6.2 \pm 6.7) \cdot 10^{-8}$	$(-1.8 \pm 1.9) \cdot 10^{-6}$	$(7.0 \pm 7.5) \cdot 10^{-20}$	$(-2.5 \pm 2.7) \cdot 10^{-20}$
Cs	$(-2.2 \pm 8.6) \cdot 10^{-6}$	$(-0.7 \pm 2.6) \cdot 10^{-6}$	$(-1.8 \pm 6.9) \cdot 10^{-19}$	$(-0.3 \pm 1.0) \cdot 10^{-20}$
^{199}Hg	$(-0.7 \pm 1.0) \cdot 10^{-12}$	$(-3.6 \pm 5.1) \cdot 10^{-13}$	$(-1.6 \pm 2.3) \cdot 10^{-26}$	$(-1.6 \pm 2.3) \cdot 10^{-25}$
^{129}Xe	$(4.5 \pm 4.7) \cdot 10^{-10}$	$(6.0 \pm 6.2) \cdot 10^{-10}$	$(-7.7 \pm 7.9) \cdot 10^{-24}$	$(-3.6 \pm 3.7) \cdot 10^{-23}$
^{171}Yb	$(2.4 \pm 1.8) \cdot 10^{-9}$	$(1.2 \pm 0.9) \cdot 10^{-9}$	$(6.0 \pm 4.6) \cdot 10^{-23}$	$(6.0 \pm 4.6) \cdot 10^{-22}$
^{225}Ra	$(2.3 \pm 3.5) \cdot 10^{-9}$	$(-5.8 \pm 8.7) \cdot 10^{-10}$	$(-0.7 \pm 1.1) \cdot 10^{-20}$	$(-3.4 \pm 5.0) \cdot 10^{-20}$
TlF	$(2.4 \pm 4.1) \cdot 10^{-11}$	$(-0.7 \pm 1.2) \cdot 10^{-9}$	$(2.7 \pm 4.6) \cdot 10^{-23}$	$(-1.0 \pm 1.6) \cdot 10^{-23}$

Table 5.4: Single-parameter ranges allowed by each of the EDM measurements given in Tab. 5.1 and coefficients from Tab. 5.3. The neutron EDM itself is best constrained by the direct experimental measurement using neutrons, see Tab. 5.1.

correlations between the leptonic and hadronic model parameters in the global analysis. The other closed-shell systems lead to much weaker limits, but will still be needed to constrain the 3-dimensional hadronic model space. Note that we are fixing the relation between the neutron and proton parameters in the Lagrangian, Eq. (2.41), but from Tab. 5.3 we know that different measurements probe different mixtures of these two parameters.

5.3 Global analysis

To combine the available EDM measurements and analyze them in terms of the hadronic-scale Lagrangian and its parameters, as in Eq. (2.42), we use the established SFitter analysis tool. It constructs a global likelihood with a comprehensive uncertainty treatment and analyses it regarding high-dimensional correlations. Depending on the preferred statistical framework, lower-dimensional and one-dimensional likelihoods for the individual model parameters can be derived by profiling or marginalization. Assuming that experimental uncertainties are Gaussian, profiling and marginalization have to lead to identical results. For the theory uncertainties, discussed in Sec. 5.3.5, the difference between the two approaches makes a formal, but not significant difference.

5.3.1 SFitter framework

SFitter as tool for global analyses was thoroughly introduced in Chapter 3. Thus we provide here only a short review of SFitter used for EDMs. It has a strong focus on the correct treatment of uncertainties, including theory uncertainties [223], and can also switch between Bayesian and frequentist approaches [75, 99], and includes some published experimental likelihoods [1].

Our first SFitter analysis of EDMs relates the 11 measurements from Tab. 5.1 to the seven model parameters in Eq. (2.42). In general, SFitter includes statistical, systematic, and theory uncertainties. The heart of SFitter is the fully exclusive likelihood as a function of model and nuisance parameters. All measurements are described as uncorrelated, with the individual systematic uncertainties described by either Poisson or Gaussian likelihood. Statistical uncertainties are, usually, uncorrelated as well and described by a Poisson distribution, turning into a Gaussian for high statistics. Experimental systematics are assumed to have a Gaussian shape, but can be described by any nuisance parameter.

For the EDM analysis, the situation is relatively simple. First, from Eq. (5.1) we know that all observables depend on the model parameters linearly. Second, we can combine the statistical and systematic experimental uncertainties into the symmetric Gaussian error bars given in Tab. 5.1. Finally, we do not have to consider nuisance parameters, if we assume that the likelihood has a Gaussian form for each independent measurement. This Gaussian assumption also implies that for uncorrelated uncertainties, a profile likelihood and Bayesian marginalization will yield the same result.

Theory uncertainties have no well-defined likelihood shape, and no maximum, but they can be thought of as a range [223]. A flat theory uncertainty is not parametrization-invariant, as one would have expected from a fixed range, but without a preferred central value we consider it conservative. With this implementation in SFitter following a flat distribution, the central value of the parameter can be shifted within this range at no cost in the likelihood. Therefore, compared to analyses without considering theory uncertainties, the impact of the central value of certain parameters is negligible for larger ranges allowed by theory uncertainties. For the EDM global analysis, theory uncertainties significantly affect most α-values: The central values given in Tab. 5.3 mainly impact implementations that neglect theory uncertainties, while the corresponding ranges are by far more relevant when theory uncertainties are taken into account.

To construct the exclusive likelihood, SFitter evaluates EDM predictions over the entire model parameter space. It uses a Markov chain to encode the likelihood in the distribution of points covering the model space. To remove nuisance parameters or to extract limits on a reduced number of model parameters, SFitter can employ a profile likelihood or a Bayesian marginalization [75, 99]. These two methods will give different results, except for the case when uncorrelated Gaussians are calculated. While profiling over flat theory uncertainties and Gaussian experimental uncertainties leads to the RFit [227] prescription, profiling over two parameters with flat likelihood leads to linearly added uncertainties even for uncorrelated parameters.

While Eq. (5.1) suggests a homogeneous set of model parameters, the typical sizes of the model parameters in Eq. (2.42) and the α-values in Tab. 5.3 can be extremely different. For numerical reasons, we internally re-scale each model parameter and each α-value such that all model parameters are evaluated with a similar size. Concretely, this means

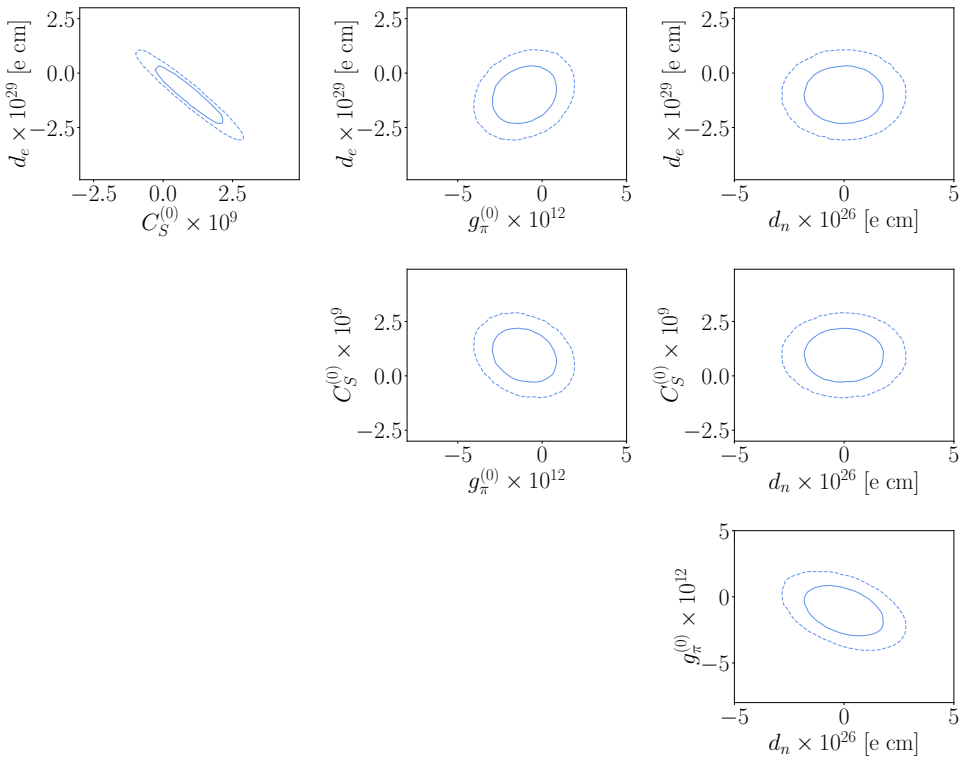

Figure 5.1: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n\}$, based on all EDM measurements but neglecting theory uncertainties. The ellipses indicate 68% and 95% CL.

rescaling d_e by a factor 10^{29}, $C_S^{(0)}$ by 10^9, $g_\pi^{(1)}$ and $g_\pi^{(0)}$ by 10^{10}, $C_T^{(0)}$ by 10^8, $C_P^{(0)}$ by 10^6, and $d_{n/p}$ by 10^{23}. These rescalings are also reflected in the way we present our results.

5.3.2 Well-constrained model sub-space

As a starting point for the global analysis and to be able to understand the main features of this newly implemented measurements and model space, we consider a sub-space of well-known and well-constrained parameters. Following the discussion in Sec. 5.2, we expect d_e and $C_S^{(0)}$ to be constrained well by the open-shell molecules HfF^+ and ThO. Similarly, the hadronic parameters $g_\pi^{(0)}$ and d_n are strongly constrained by the neutron and Hg EDMs. This means the model sub-space

$$\left\{ d_e, C_S, g_\pi^{(0)}, d_n \right\} \tag{5.15}$$

should be constrained well by the full set of measurements given in Tab. 5.1.

More than in the numerical constraints and boundaries, we are interested in the correlation of the constraints for these four model parameters. In Fig. 5.1 we show these correlations extracted as 2-dimensional profile likelihoods from the fully exclusive, 4-dimensional likelihood. Three structural aspects stick out: (i) A strong anti-correlation between d_e and $C_S^{(0)}$; (ii) a very slight anti-correlation between $g_\pi^{(0)}$ and d_n; and (iii) essentially no correlations between the $\{d_e, C_S^{(0)}\}$ and $\{g_\pi^{(0)}, d_n\}$ parameter subsets.

Within the leptonic sector, the strong correlations between d_e and $C_S^{(0)}$ and their independence from the remaining parameter space are expected to remain for the full global analysis. The reason is that it is induced by the strongest measurements of HfF^+ and ThO, and according to our parameterization as shown in Tabs. 5.3 and 5.4, those two measurements are not affected by any other model parameter. This means the upper-left panel of Fig. 5.1 factorizes from our global EDM analysis, and we can consider the remaining model parameters separately and without including the HfF^+ and ThO measurements further.

Moving from the leptonic to the hadronic sector where the situation is different for the model parameter. Similar to the ThO and HfF^+ measurements, the neutron and Hg measurements are three orders more constraining than the other measurements. However, they are constraining more model parameters than just d_n and $g_\pi^{(0)}$, so for the sub-space analysis we could have as well chosen $g_\pi^{(0)}$ vs $g_\pi^{(1)}$, without any change in the conclusion. This leads to the expansion of the hadronic parameter space as the next step to see what patterns might emerge. The parameters d_e and $C_S^{(0)}$ will be kept factorized for the rest of our global analysis.

To understand the role of approximately flat directions in our model space, we can diagonalize and invert the α-matrix given in Tab. 5.3 for the well-constrained subsystem. To be able to invert the α-matrix, we have to truncate it to a squared form. We know that the leading constraints in the $d_e - C_S^{(0)}$ plane come from the ThO and HfF^+ measurements and can be described by the invertible relation

$$\begin{pmatrix} d_{\mathrm{HfF}^+} \\ d_{\mathrm{ThO}} \end{pmatrix} = \begin{pmatrix} \alpha_{\mathrm{HfF}^+,d_e} & \alpha_{\mathrm{HfF}^+,C_S^{(0)}} \\ \alpha_{\mathrm{ThO},d_e} & \alpha_{\mathrm{ThO},C_S^{(0)}} \end{pmatrix} \begin{pmatrix} d_e \\ C_S^{(0)} \end{pmatrix} . \tag{5.16}$$

We can diagonalize the α-submatrix, find the eigenvalues 1.0 and $5.93 \cdot 10^{-21}$, and invert it to give the model parameters as a function of the measurements,

$$\begin{aligned} \begin{pmatrix} d_e \\ C_S^{(0)} \end{pmatrix} &= \begin{pmatrix} 2.55 & -1.55 \\ -1.69 \cdot 10^{20} & 1.69 \cdot 10^{20} \end{pmatrix} \begin{pmatrix} d_{\mathrm{HfF}^+} \\ d_{\mathrm{ThO}} \end{pmatrix} \\ &= \begin{pmatrix} 2.55 d_{\mathrm{HfF}^+} - 1.55 d_{\mathrm{ThO}} \\ (-1.68 d_{\mathrm{HfF}^+} + 1.69 d_{\mathrm{ThO}}) \cdot 10^{20} \end{pmatrix} . \end{aligned} \tag{5.17}$$

Approximately flat directions in model space appear because the measurements are uncorrelated. We could determine d_e much more precisely and without any effect from $C_S^{(0)}$ if we could measure the fully correlated combination $(2.55 d_{\mathrm{HfF}^+} - 1.55 d_{\mathrm{ThO}})$, which unfortunately is not possible.

Deliberately correlated measurements can be envisioned, for instance via comagnetometry, as already used for the neutron and Xe measurements. In this interpretation it could be advantageous to perform correlated measurements of two systems, with comparable sensitivity to both system EDMs; unlike the usual comagnetometer implementation, where it is usually assumed that one can be neglected entirely.

5.3.3 Hadronic parameters from closed-shell systems

For the purely hadronic sector, we define a second simplified model parameter space,

$$\left\{ g_\pi^{(0)}, g_\pi^{(1)}, d_n \right\} . \tag{5.18}$$

All three parameters are constrained by the neutron and closed-shell EDMs, while we know from the above discussion that the constraints from closed-shell systems on d_e and $C_S^{(0)}$ are weaker than those from their open-shell counterparts.

From Tab. 5.4 we see that the neutron and Hg measurements strongly constrain two of the three hadronic model parameters in Eq. (5.18). To understand the correlations structure induced by the remaining closed-shell measurements, we show the correlated constraints of the six possible parameter pairs for the closed-shell measurements on the different 2-dimensional sub-spaces of Eq. (5.18).

Starting with the left panel of Fig. 5.2, with the $g_\pi^{(0)}$ vs. $g_\pi^{(1)}$ plane, we can see different correlation patterns. This implies that performing a global analysis will constrain this sub-space much better than any single pair of measurements. In the center panel, the situation changes when we look at the correlations with the neutron EDM-parameter d_n. Three combinations are aligned to similar negative correlations between d_n and $g_\pi^{(0)}$. In the right panel, the situation is similar for $g_\pi^{(1)}$, with a positive correlation and less striking. In both cases, the exceptions are the combinations of Xe with Ra, TlF, and Yb, which constrain d_n extremely well and without any correlation with $g_\pi^{(0,1)}$. While this sounds like an advantage, we remind ourselves that from Tab. 5.4 and Fig. 5.1 we know that those limits on d_n are three orders of magnitude weaker than what can be expected from including the neutron and the Hg measurements.

Altogether, Fig. 5.2 confirms that the constraints of the sub-leading four closed-shell measurements on the 3-dimensional hadronic parameter space of Eq. (5.18) are correlated in a non-trivial manner. Evaluating these correlations requires a global analysis of the formally over-constraining set of closed-shell measurements.

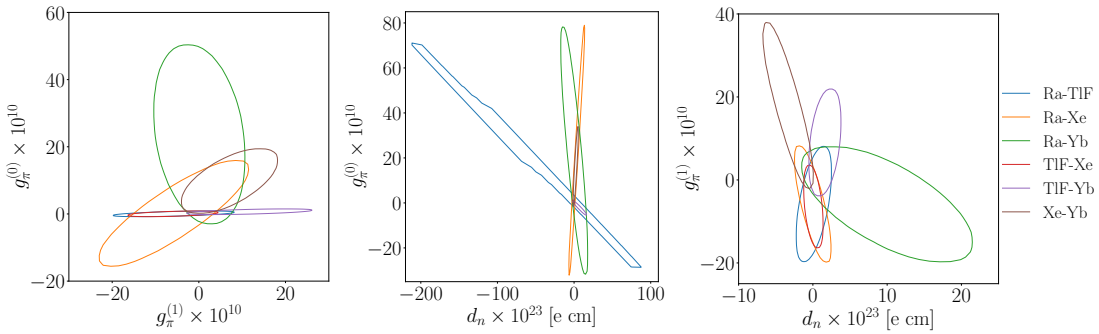

Figure 5.2: Correlations from three 2-dimensional analyses in the $\{g_\pi^{(0)}, g_\pi^{(1)}, d_n\}$ parameter space, each based on a different pair of closed-shell EDM measurements, as indicated by the color. The ellipses indicate 68% CL, neglecting theory uncertainties.

5.3.4 Poorly constrained model parameters

Finally, we combine the effects of all remaining parameters,

$$\left\{ C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n \right\}, \tag{5.19}$$

ignoring the ThO and HfF^+ measurements, which constrain the factorized $d_e - C_S^{(0)}$ subspace, and also ignoring the neutron and Hg measurements. The latter constrain the above parameters, but because they are much stronger than all other measurements, they will induce narrow correlations in the allowed 5-dimensional parameter space, which will make it hard to constrain other parameter combinations.

Narrow correlations in a higher-dimensional parameter space vanish when we profile the likelihood onto 2-dimensional correlations or even single model parameters. For example, consider a 2-dimensional parameter space constrained by one strong and one weak measurement. The strong measurement leads to a narrow correlation between the two parameters. When we extract the profile likelihood for one model parameter, we can adjust the other model parameter so that the two parameters trace this narrow correlation. This way, the entire length of the correlation pattern gets projected onto the

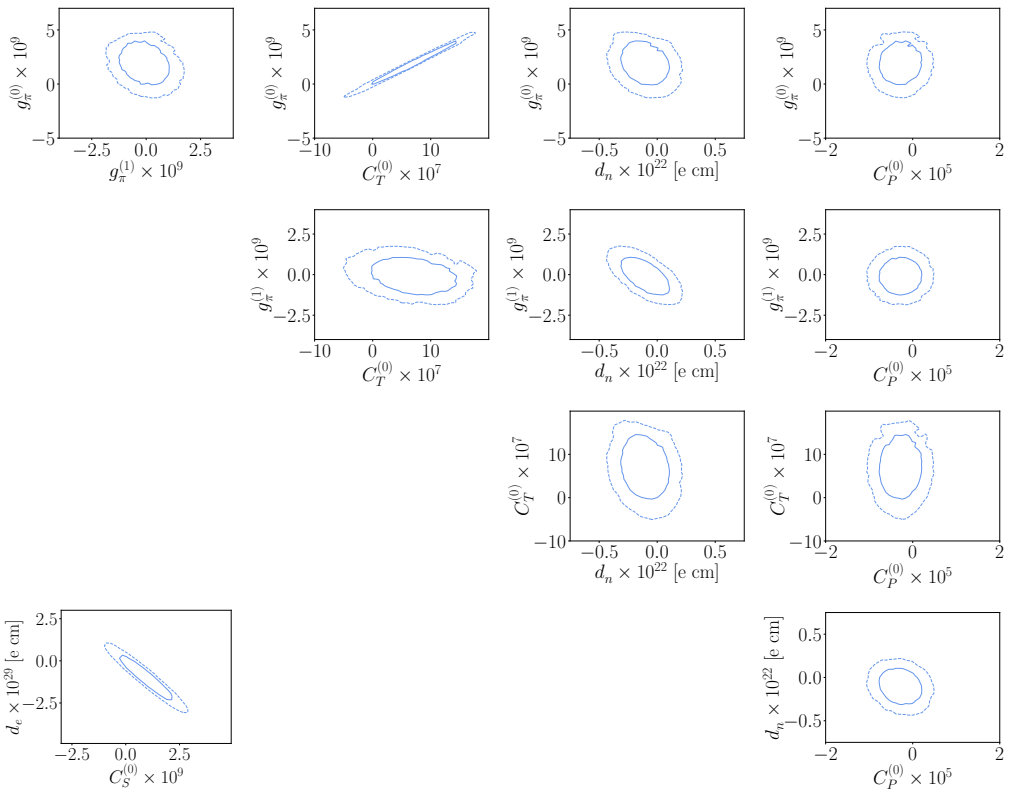

Figure 5.3: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n\}$ and the factorized $d_e - C_S^{(0)}$ plane from Fig. 5.1. We ignore the neutron and Hg measurements, which induce narrow correlation patterns in the 5-dimensional parameters space and do not affect the profiled 2-dimensional correlations. The ellipses indicate 68% and 95% CL, neglecting theory uncertainties.

1-dimensional profile likelihood. The weak measurement dominates the individual profile likelihoods, while the strong measurement allows us to link the second model parameter from a first model parameter precisely.

In our case of the hadronic sub-space, this implies that as long as the correlations induced by the neutron and Hg measurements cross the entire parameter space, we can ignore these two measurements and their induced correlation patterns in the following discussion of 2-dimensional correlations and single-parameter profile likelihoods. Losing the best few measurements contributing to the global analysis is an unfortunate effect of the conservative profile likelihood approach, but it should also exist for standard error ellipses using Bayesian marginalization. The main difference is that this kind of effect is numerically extremely challenging to compute using marginalization, while it is, essentially, trivial for the profile likelihood approach.

Given these considerations of correlations, we are left with five model parameters, constrained by seven measurements of comparable constraining power. The only exception we see in Fig. 5.3 is a strong leading correlation between $g_\pi^{(0)}$ and $C_T^{(0)}$, induced by the fact that the TlF measurement is leading for both parameters by one order of magnitude compared to others. All other model parameters are nicely constrained. The allowed range, for instance for $g_\pi^{(0)}$ is of the order 10^{-9}. This can be compared to the constraints from Fig. 5.1, of the order 10^{-12}. The same hierarchy of measurements can be observed for $g_\pi^{(1)}$ and for d_n, as confirmed by Tab. 5.4. This means that the 5-dimensional allowed parameter space illustrated by Fig. 5.3 is crossed by two correlation patterns, roughly three orders of magnitude more narrow than the full parameter space. We emphasize that explaining this extremely narrow correlation poses a numerical fine-tuning problem in the model parameter space, which the profile likelihood does not address.

As before, we can truncate the number of available measurements for the poorly constrained 5-dimensional subspace given in Eq. (5.19) to the five leading measurements, and invert the corresponding α-submatrix to find

$$\begin{pmatrix} C_T^{(0)} \\ C_P^{(0)} \\ g_\pi^{(0)} \\ g_\pi^{(1)} \\ d_n \end{pmatrix} = \begin{pmatrix} 7.14 \cdot 10^{18} & -3.21 \cdot 10^{18} & 6.50 \cdot 10^{19} & -8.77 \cdot 10^{13} & -1.94 \cdot 10^{15} \\ -1.03 \cdot 10^{17} & -5.40 \cdot 10^{19} & 1.09 \cdot 10^{21} & -3.74 \cdot 10^{14} & -3.26 \cdot 10^{16} \\ -2.49 \cdot 10^{14} & -1.45 \cdot 10^{17} & 2.83 \cdot 10^{18} & -2.37 \cdot 10^{12} & -8.60 \cdot 10^{13} \\ 3.10 \cdot 10^{14} & -3.46 \cdot 10^{16} & -2.56 \cdot 10^{18} & 1.58 \cdot 10^{12} & 5.49 \cdot 10^{13} \\ 0. & 0. & 0. & 0. & 1. \end{pmatrix} \begin{pmatrix} d_{\mathrm{Tl}} \\ d_{\mathrm{Hg}} \\ d_{\mathrm{Xe}} \\ d_{\mathrm{TlF}} \\ d_{\mathrm{n}} \end{pmatrix}. \tag{5.20}$$

5.3.5 Theory uncertainties

Theory uncertainties always appear when we use a quantum field theory to predict observables, like EDMs, from Lagrangian parameters: No calculation method is arbitrarily precise, and a variety of systematic errors can affect the accuracy. While there is some hope in estimating and controlling uncertainties for small expansion parameters, like in a high-energy EFT expansion, it is far more difficult for uncertainties associated with QCD observables at low energies (whether from lattice calculations or sum-rule estimates). Also, quantifying the precision of nuclear physics calculations and linking them to an effective quantum field theory is a challenge in itself. On the other hand, we have to estimate all of these uncertainties, and we can only ignore them after having shown

that they are significantly smaller than the experimental uncertainties of the associated measurements.

For our global EDM analysis, the theory uncertainties we have to consider only affect the α_{i,c_j} in Eq. (5.1). The estimated range of the theory uncertainties for every parameter is also given in Tab. 5.3. Currently, we treat them in SFitter as uncorrelated theory uncertainties, however, this assumption can be modified if necessary. Because of the flat likelihood as a function of the theory nuisance parameter, the profile likelihood approach leads to the theory uncertainties adding linearly, weighted by the respective model parameter. By profiling over the independent α-ranges, the numerical evaluation gets simplified in two ways: First, any parameter–observable pair for which α is compatible with zero will effectively be removed from the global analysis, because the optimal choice of α will remove all contributions from the corresponding model parameter; second, even if we cannot choose α such that measurement and prediction agree, we can choose it to maximize the likelihood and to minimize the impact of the measurement, which means we choose the smallest allowed absolute value of α.

Same as in Sec. 5.3.2 we start with the well-constrained 4-dimensional subspace $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n\}$. Again, d_e and $C_S^{(0)}$ are constrained by the open-shell molecules HfF^+ and ThO, just as without theory uncertainties. From Tab. 5.3 we see that we can ignore the theory uncertainty in relating the electron EDM parameter d_e to these systems. In the hadronic sector, the effect of theory uncertainties on the relation of d_n and $g_\pi^{(0)}$ originating from the Hg and neutron measurement is either trivial or reasonably small,

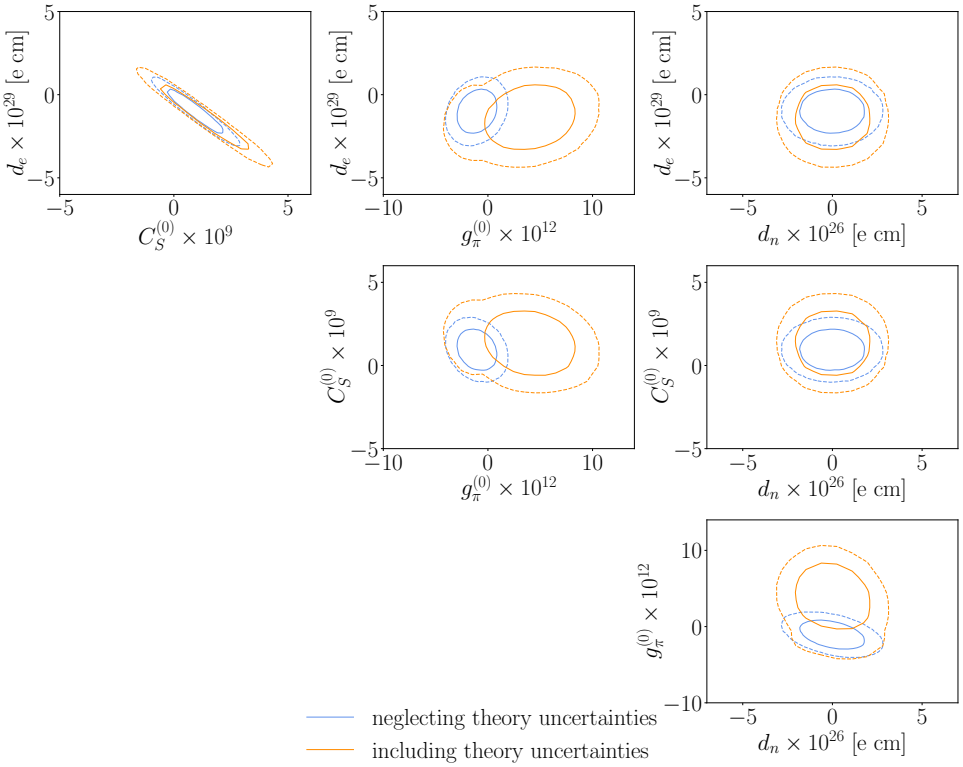

Figure 5.4: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n\}$. The orange curves show the effect of theory uncertainties on the results of Fig. 5.1. The ellipses indicate 68% and 95% CL.

although for the neutron, this situation should not be over-interpreted. The quoted uncertainty arises from propagating ranges for the constants within Eq. (5.5), and not from careful evaluations of the chiral expansion itself.

In Fig. 5.4, we show the numerical impact of the theory uncertainties on the 2-dimensional correlations. The slightly stronger HfF^+ measurement, which determines the width of the correlation pattern, is only minimally affected by the theory uncertainties, while the larger theory uncertainties on the ThO measurement extend the length of the ellipse visibly. The main effect of the theory uncertainties is on $g_\pi^{(0)}$, where they shift the allowed ranges from slightly negative to sizeable positive values. This analysis includes all measurements, so according to Tab. 5.4, $g_\pi^{(0)}$ is most strongly constrained by the Hg measurement (with a negative central value) and TlF (with a large and positive central value). Adding the theory uncertainties weakens the Hg measurement, which means the two leading constraints get balanced differently, and the entire range moves to the positive values preferred by the TlF measurement.

Moving on, we can now look at the effect of the theory uncertainties on the less-constrained hadronic sector for closed-shell systems, discussed in Sec. 5.3.3. Here, Tab. 5.3 shows sizeable, order-one theory uncertainties. In addition, some of the α-values include an allowed zero value when we include theory uncertainties. Specifically, $g_\pi^{(1)}$ will no longer be constrained by the Yb measurement, and it will also lose the Hg constraint. Also,

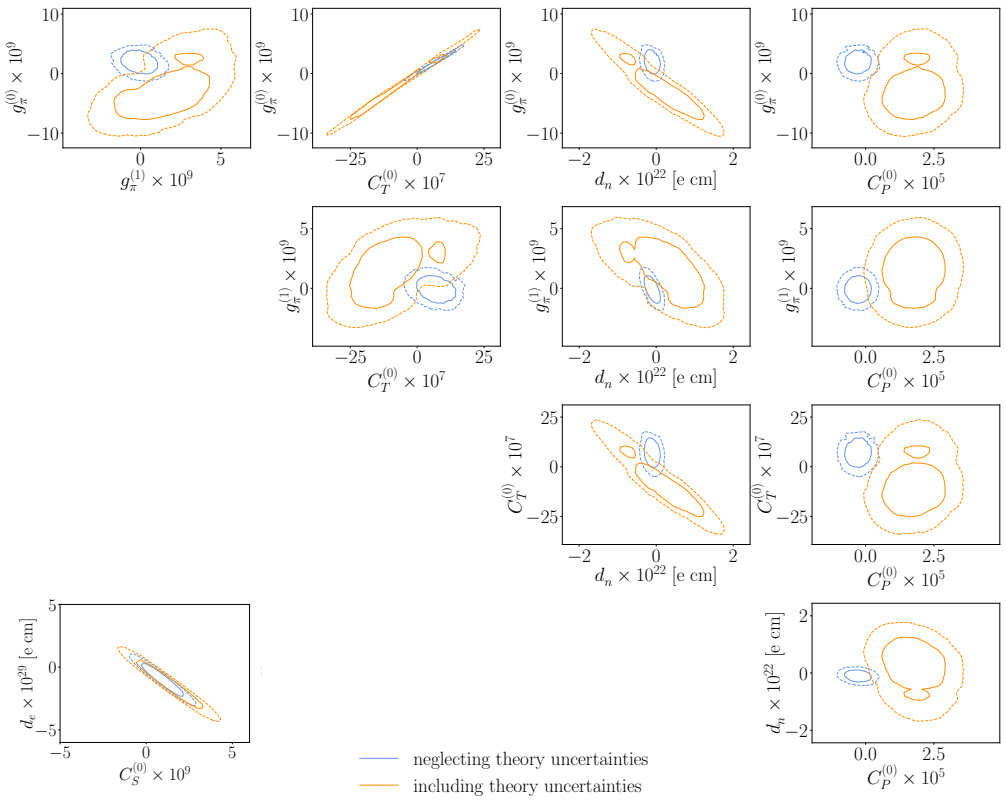

Figure 5.5: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n\}$, and the factorized $d_e - C_S^{(0)}$ plane from Fig. 5.4. The orange curves show the effect of theory uncertainties on the results of Fig 5.3. The ellipses indicate 68% and 95% CL.

some of the theory uncertainties entering the closed-shell sector are not consistent with $\alpha = 0$ but are large, so we expect a sizeable impact on the global analysis.

Finally, we can look at all seven EDM parameters. As for the case without theory uncertainties, the neutron and Hg limits are much more constraining than the other measurements of the hadronic sector. Following the argument given in Sec. 5.3.4, this means that 2-dimensional correlations and single-parameter limits extracted by profiling over the likelihood will not be impacted by these strong measurements. The shift in the constrained 2-dimensional correlations is shown in Fig. 5.5. In comparison, their effect on the factorized parameters d_e and $C_S^{(0)}$ (copied from Fig. 5.4) is mild, and the constraints on the hadronic model parameters are significantly weaker. The leading correlation in this parameter extraction we found out to come from the TlF measurement constraining $g_\pi^{(0)}$ and $C_T^{(0)}$ in a correlated manner. When we allow for the additional theory uncertainties, this correlation expands almost to a flat direction. In turn, this large effect extends to the entire hadronic sector, weakening and shifting essentially all constraints. The large shifts addressed in Fig. 5.5, mostly seen for $C_P^{(0)}$, originate in the treatment and implementation of theory uncertainties. The dominant measurement for $C_P^{(0)}$ suffers from large theory uncertainties, allowing for a shift in estimating the central parameter value, as discussed in Sec. 5.3.1. This clearly emphasizes the impact of theory uncertainties on the correlations and allowed ranges of the model parameters. To obtain a more coherent picture of these parameter correlations and ranges, theory uncertainties must be reduced (ideally including also a more nuanced treatment of the corresponding likelihoods). There are already some measurements and parameter combinations, like the d_e and $C_S^{(0)}$ subspace, for which theory uncertainties are reasonably well under control (as compared to other elements of Tab. 5.3). For this case, theory uncertainties produce no significant change of the correlation behavior.

5.4 Outlook

EDMs are extremely sensitive, targeted probes for one of the most important symmetries of elementary particles, which is directly related to the baryon asymmetry in the universe. The number of EDM measurements performed in various systems has grown rapidly in recent years. This leads to the question of how the different measurements are contributing to constraining and understanding CP violation in terms of a fundamental Lagrangian.

We can choose different Lagrangians to answer this question, starting with UV-complete models versus EFTs. Without a specific hint for BSM physics, we chose an EFT description. Next, we have a choice of different energy scales with different degrees of freedom. For our first SFitter analysis, we rely on the hadronic-scale Lagrangian, valid at the GeV scale and describing the interactions of electrons and nucleons. After relating the hadronic-scale Lagrangian to its weak-scale SMEFT counterpart we want to constrain the seven Lagrangian parameters given in Fig. 5.6 through 11 independent measurements given in Tab. 5.1.

As a toy analysis, we examine the single-parameter constraints from all individual EDM measurements. These limits are all driven by the same small set of highly constraining measurements, such as the open-shell molecular ion HfF^+, the neutron EDM, or the closed-shell atom Hg. The extremely strong constraints indicated in Fig. 5.6 do not allow for a cancellation of contributions from two model parameters to a given measurement at the price of creating a numerical fine-tuning problem in the model parameter space.

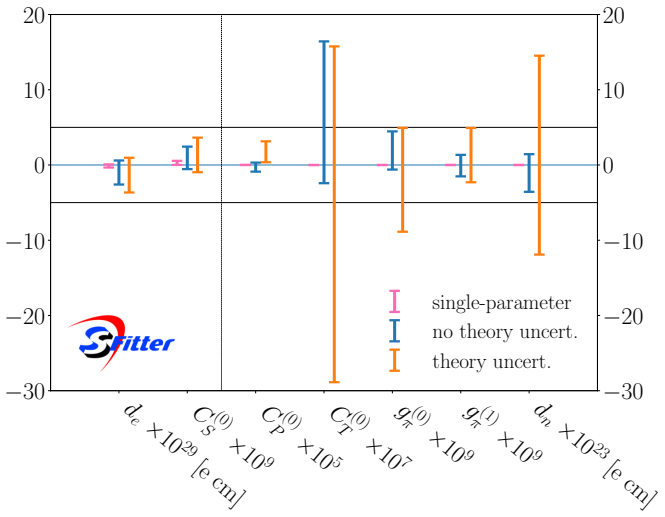

Figure 5.6: 68% CL constraints from the global EDM analysis on the parameters of the hadronic-scale Lagrangian. We show (i) hugely over-constrained single-parameter ranges allowed by the best available measurement; (ii) over-optimistic allowed ranges for profiled single parameters, ignoring theory uncertainties; (iii) allowed ranges for profiled single parameters, including experimental and theory uncertainties.

While the single-parameter estimates indicate the strength of an experiment looking for a sign of CP violation, they should not be confused with a measurement of a given parameter.

For our global analysis, we use SFitter, with its focus on statistical interpretation and a comprehensive uncertainty treatment. First, we ignore all theory uncertainties and only consider experimental uncertainties as uncorrelated, following a Gaussian distribution (combining the statistical and systematic uncertainties reported in the respective papers). In this case, Bayesian marginalization and profile likelihood yield the same result.

In Sec. 5.3.2, we find that a small set of powerful measurements constrains the electron-hadron interactions as well as a subset of the hadronic sector. The correlated limits on the electron EDM parameter d_e and the scalar coupling $C_S^{(0)}$ are especially strong and factorize from the hadronic sector. Next, we find in Sec. 5.3.3 that the constraints on the hadronic sector from the closed-shell systems show rich correlations, motivating our global analysis further, as described in Sec. 5.3.4. For the hadronic parameters, the narrow correlations from the strong neutron and Hg constraints do not appear in profiled 2-dimensional correlations or single-parameter limits. As a result, the hadronic model parameters are constrained much worse than the single-parameter results suggest, as shown in Fig. 5.6.

Finally, we show the same limits but include theory uncertainties. Such uncertainties always appear when we relate measurements to fundamental Lagrangian parameters. We assume a flat likelihood within allowed ranges of the factors relating the Lagrangian parameters to the EDM predictions. While the impact of the theory uncertainties on the factorized $\{d_e, C_S^{(0)}\}$ sector is relatively mild, the correlated analysis of the hadronic parameters leads to a significant weakening of the constraints on all model parameters.

Chapter 6

Machine Learning basics

This chapter serves as a short introduction to machine learning (ML) and, more specifically, deep learning. One of the many applications for ML networks is regression. Such an application will be discussed later in the scope of precision and uncertainty estimation in amplitude regression. Thus, this chapter is structured in the following way: In Sec. 6.1, we discuss the basic concepts of fully connected networks, their loss functions, and their optimization. Since every measurement and prediction has to come with uncertainties, we need network architectures that can capture these uncertainties. Therefore, the different uncertainty types entering ML applications are introduced in Sec. 6.2 and later implemented into actual network architectures in Sec. 6.3.

6.1 Introduction to neural networks

This section provides a brief introduction to neural networks (NN). We start by discussing the setup of a simple architecture consisting of only linear layers with some non-linearities in between. Next, we explain the construction of a loss function and network optimization.

6.1.1 Fully connected networks

The core task of NN architectures is approximating high-dimensional, complex functions using a series of more straightforward, linear functions with non-linearities in between. These non-linearities are the so-called activation functions. This links every input feature to the output features, which is then called a fully connected network.

A single layer of this network is constructed as

$$y = \Phi(Wx + b) \,, \tag{6.1}$$

with Φ the activation function, x as input parameters with D_x dimensions and y the output parameters with D_y dimensions. b is the so-called bias, having D_y dimensions and W is a $D_y \times D_x$ dimensional matrix. In this layer, W and b are the trainable parameters of the network, and they are optimized with respect to the corresponding training objective. This most simple type of network architecture, consisting of linear layers followed by non-linearities or activation functions, is also called a Multilayer perceptron (MLP)

network. Some popular choices as activation functions include these component-wise functions such as [274, 275]

$$\mathrm{ReLU}(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases} \tag{6.2}$$

$$\mathrm{LeakyReLU}(x) = \begin{cases} \beta x & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases} \tag{6.3}$$

$$\mathrm{GELU}(x) = x \cdot \frac{1}{2}[1 + \mathrm{erf}(x/\sqrt{2})] \tag{6.4}$$

$$\mathrm{Sigmoid}(x) = \frac{1}{1 + \exp(-x)} \, . \tag{6.5}$$

The factor β in the LeakyReLU function ensures a small slope for negative values of x instead of a flat slope as in the ReLU case. The corresponding β is determined before the training process starts and thus not modified during training. Often, no activation function is applied to the last layer of a network if the output should be unconstrained. In the considered example of a regression task for amplitude surrogates, we are interested in the unconstrained output and compare it to the truth labels to determine the precision of the results.

6.1.2 Loss function and optimization

Optimizing a neural network requires a loss function that connects all the trainable parameters with the training data D. This loss function aims to represent the posterior $p(\theta|D)$, which we want to maximize as a training objective. Since the posterior is not accessible in the training process, equivalent to maximizing the posterior, we can maximize the likelihood function, which is given as

$$\mathcal{L} = -\log p(D|\theta) = -\sum_{i=1}^{N} \log p(D_i|\theta) \, . \tag{6.6}$$

For simplicity reasons, we can replace the likelihood with the log-likelihood to treat constant prefactors as additional, constant terms in the loss. Applying this likelihood approach to a regression task, we assume a Gaussian distribution as an ansatz with no uncertainties for further simplifications. This ansatz results in the mean squared error (MSE) as a loss function, where the network f_θ with its trainable network parameter θ connects the input x_i with the true value/truth label y_i. It can be written as

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} (y_i - f_\theta(x_i))^2 \, . \tag{6.7}$$

Instead of maximizing the likelihood or log-likelihood, we can equivalently minimize the negative log-likelihood function as loss. To minimize this negative log-likelihood function, we apply the gradient descent method as a computationally efficient approach, where the trainable parameters are updated in small steps in the opposite direction of the gradient, the direction of the steepest descent, to minimize them. This method is performed by calculating the gradient of the loss function with respect to the trainable parameters,

$$\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{L} \, , \tag{6.8}$$

where α describes the learning rate. A downside of this method is that it can easily get stuck in local minima when performed on the entire training data. To prevent the network from getting stuck, the training dataset is randomly divided into smaller batches, and then the optimization is performed batch-wise. This adapted method is called stochastic gradient descent (SGD). Also, more sophisticated algorithms are based on the SGD and implement further improvements. The Adam optimizer [276] is such an example of an improved SGD. It adapts the learning rate α based on the mean and variance of the gradient and computes the next update step based on these two parameters and not only on the raw gradients, as SGD does. Thus, Adam leads to a faster and more stable convergence. Another modification, based on Adam, is AdamW [277], which decouples the weight decay from the learning rate. Weight decay is a commonly used regularization technique to prevent overfitting. It adds a penalty term to the loss function, encouraging the network to learn smaller parameter values and pushing them toward zero. In this thesis, AdamW is used for all networks discussed.

The learning rate α can either be a fixed parameter throughout the training or be adapted with a scheduler. This has the advantage of modifying the learning rate during the training process because different learning rates are more practical at different stages of the training. At the beginning of the training, a large learning rate is preferred, leading to faster convergence, and the network is less likely to converge to a local minimum instead of the global minimum. At the end of the training process, a smaller learning rate leads to more stability in the training and increases the precision of the prediction.

There are different forms of loss functions, the one used in regression tasks being the likelihood loss. One example of such a likelihood loss is the MSE loss written down in Eq. (6.7). This MSE loss is used when we do not encounter any uncertainties. If we assume different uncertainties from the phase space, we have to modify this loss function accordingly by adding the variance σ. This updated probability function can be written as

$$p_\theta(x, y) = p_\theta(y|x)p(x) \propto \exp\left(\frac{(y - f_\theta(x))^2}{2\sigma^2}\right) p(x) \,. \tag{6.9}$$

The resulting loss function from the updated probability function is called a heteroscedastic loss. This heteroscedastic loss now includes two adaptable functions over the input space. These functions are characterized by the trainable network parameters $f_\theta(x)$ and $\sigma(x)$,

$$\mathcal{L}_{\text{het}} = \left\langle \frac{|y - f_\theta(x)|^2}{2\sigma(x)^2} + \log \sigma(x) \right\rangle_{x \sim D} \,. \tag{6.10}$$

Here, $\sigma(x)^2$ is unknown and must be learned by the network. Thus, we are able to extend the simple MSE loss to capture uncertainties from the network training.

6.2 Different types of uncertainties in ML

One important aspect of statistics is that every measurement or prediction has to come with uncertainties; otherwise, it is not complete. Measuring or predicting anything without a corresponding uncertainty for the assumed model or experimental setting is impossible. Therefore, the uncertainty treatment we know from other applications in physics, such as global analyses, has to be translated into deep learning.

With the already introduced uncertainty σ in the heteroscedastic loss, we show that it is possible for networks also to predict these uncertainties. Furthermore, the uncertainties in machine learning can be divided into two categories:

1. **Systematic or epistemic uncertainties**
   These are intrinsic uncertainties in the data, like noise or other effects, namely those originating from the experimental setup. Even after perfect training, these uncertainties do not vanish but reach a plateau. Also, they cannot be reduced with better network expressivity or improved training. Additionally, some uncertainties can arise from the network architecture, such as poorly tuned hyperparameters and network expressivity or not fully converging training.
2. **Statistical or aleatoric uncertainties**
   Here, the uncertainties originate from the limited amount of training data the network is trained on. Even if the network is expressive enough and perfectly fine-tuned, the limitations in training data lead to an uncertainty of the network weights. However, they vanish within the limits of infinite training data and perfect training.

The terms systematic and statistical uncertainties are characterized from a physicist's point of view, while most computer scientists will call them epistemic and aleatoric uncertainties [54]. In this thesis, the uncertainties are categorized as follows: We will use the terms systematic and statistical uncertainties, where the systematic uncertainties encounter noisy data or labels and structural uncertainties, like limited network expressivity, a non-optimal architecture, or choice of hyperparameters. In contrast, the statistical uncertainties arise from limited training data.

6.3 Different network architectures

This section discusses different network architectures and how they can be used in regression tasks. First, we start with Bayesian Neural Networks (BNNs) and then discuss Repulsive Ensembles (REs) as a complementary approach. Lastly, we look at Kolmogorov-Arnold Networks (KANs) and how they can help improve the accuracy of network predictions.

6.3.1 Bayesian neural networks

A BNN [52–55] can be constructed for any given model based on a likelihood loss with a fixed training data set, like in a typical regression task setting [49]. We start by considering a training dataset D with N points (x_i, y_i) and Eq. (6.6) as the loss function we want to minimize.

To make this network Bayesian, we want to sample the posterior distribution $p(\theta|D)$ of the network weights θ. Therefore, we can either directly calculate it, which is intractable, or we approximate it using a variational inference (VI) [278, 279] approach. When choosing the VI approach, the true posterior is approximated by a more straightforward, tractable distribution $q_\phi(\theta)$, with trainable parameters ϕ. The new training objective has now

become to minimize the Kullback-Leibler (KL) divergence [280] between the approximate and true posterior

$$\phi_{\text{opt}} = \arg\min_{\phi} \text{KL}(q_\phi(\theta), p(\theta|D)) \ . \tag{6.11}$$

The appearing KL divergence can be expressed in terms of the prior and using Bayes' theorem

$$\begin{aligned}
\text{KL}(q_\phi(\theta)|p(\theta|D)) &= \int \text{d}\theta \ q_\phi(\theta) \log \frac{q_\phi(\theta)}{p(\theta|D)} \\
&= \int \text{d}\theta \ q_\phi(\theta) \log \frac{q_\phi(\theta) p(D)}{p(D|\theta) p(\theta)} \\
&= -\int \text{d}\theta \ q_\phi(\theta) \log p(D|\theta) + \int \text{d}\theta \ q_\phi(\theta) \log \frac{q_\phi(\theta)}{p(\theta)} + \int \text{d}\theta \ q_\phi(\theta) \log p(D) \\
&= -\langle \log p(D|\theta) \rangle_{\theta \sim q_\phi(\theta)} + \text{KL}(q_\phi(\theta), p(\theta)) + \log p(D) \ .
\end{aligned} \tag{6.12}$$

From there on, the full log-likelihood can be written as the sum of the single log-likelihood terms obtained from independent data points,

$$\log p(D|\theta) = \sum_{i=0}^{N} \log p(y_i|x_i, \theta) \ . \tag{6.13}$$

Now, the adapted formula for a split-up training dataset into M batches is given by

$$\log p(D|\theta) \approx \frac{N}{M} \sum_{i=0}^{M} \log p(y_i|x_i, \theta) \ . \tag{6.14}$$

Since the evidence $p(D)$ does not depend on the trainable weights ϕ, it can be neglected in the loss function. With this, we can approximately calculate the expectation value of the likelihood with respect to the parameters θ by drawing a sample of $q_\phi(\theta)$ for every batch. The complete loss function reads

$$\mathcal{L}_{\text{BNN}} = -\frac{1}{M} \sum_{i=0}^{M} \log p(y_i|x_i, \theta) + \frac{1}{N} \text{KL}(q_\phi(\theta), p(\theta)) \qquad \text{with } \theta \sim q_\phi(\theta) \ , \tag{6.15}$$

where we divided Eq. (6.12) by N to obtain the log-likelihood loss term in the first term of the equation, and y_i are the truth labels of the data set D. The second term acts as a regularization of the network weights.

We choose uncorrelated Gaussians for the prior and posterior distributions. While these are restrictive assumptions on both, the network can still model complex uncertainties. The reason for that lies in the structure of the network, with many linear layers and non-linearities in between. Therefore, an injection of a Gaussian approximation in early layers can lead to non-linear effects on the output [54].

The prior is often chosen as

$$p(\theta) \sim \mathcal{N}(0, \sigma_p) \ , \tag{6.16}$$

with $\mu_p = 0$ and standard deviation σ_p. The posterior is approximated using factorized

Gaussians

$$q_\phi(\theta) = \prod_j^K \mathcal{N}(\theta_j|\mu_j, \sigma_j) \ , \tag{6.17}$$

where $\phi = \{\mu_j, \sigma_j\}$ and K is the number of parameters in the neural network. Comparing the parameter numbers used in the training process for an MLP with a BNN, the MLP has θ_j as trainable parameters, while the BNN has parameter pairs $\phi_j = (\mu_j, \sigma_j)$. Thus, the number of parameters in the training process doubles for a BNN compared to an MLP. With the Gaussian distribution as a choice for the prior and posterior, the KL divergence can be computed analytically

$$\mathrm{KL}(q_\phi(\theta), p(\theta)) = \sum_j \left(\log \frac{\sigma_p}{\sigma_j} + \frac{\sigma_j^2 + \mu_j^2}{2\sigma_p^2} - \frac{1}{2} \right) \ . \tag{6.18}$$

Sampling from $q_\phi(\theta)$ while computing the gradients is possible by using the reparameterization trick [281]

$$\theta = \mu + \sigma\epsilon \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, 1) \ . \tag{6.19}$$

Since the gradients $\partial\theta/\partial\mu$ and $\partial\theta/\partial\sigma$ can be computed, we can use backpropagation to compute the gradients of the loss with respect to μ and σ.

6.3.2 Repulsive ensemble

An ensemble of networks can be used as alternative to BNNs for uncertainty estimation. However, the ensemble members have to explore the full space of local minima. Furthermore, it has to be ensured that the estimated uncertainty covers the entire probability distribution [282, 283]. Using a network ensemble is another way to go beyond a single set of best-fit parameters.

Starting from an ensemble of networks with different initializations, the training results in a range of possible training outcomes by converging to different local minima in the phase space. Based on Eq. (6.8), we have seen an update of network parameters using the gradient descent method, where only a single network was updated. However, in an ensemble of networks, all members must be updated simultaneously when minimizing the loss function. As a general form for updating the weights, we obtain

$$\theta^{t+1} = \theta^t + \alpha \nabla_{\theta^t} \log(p(\theta^t|D) \ , \tag{6.20}$$

with the learning rate α and the training data set D. The shown update rule does not lead to or even require interaction between members. Thus, several of them can converge to the same minimum, which can cause the posterior to collapse and prevent the ensemble from providing a reasonable uncertainty estimation. On the other hand, if the ensemble is sufficiently spread out and the members approach different local minima, it can give us a reasonable distribution over parameter space, thus complementing BNNs in deriving the loss function locally and globally. Therefore, we insert an interaction between the ensemble members by introducing a repulsive kernel $k(\theta, \theta_j)$, turning them into a RE. This additional term forces the ensemble to spread around the (local) minimum in the loss. Also, the repulsive kernel $k(\theta, \theta_j)$ should increase with decreasing distance of one ensemble member θ to all other members. By connecting the different members via a

kernel function and ensuring the spread over the whole phase space, REs can provide decent results, even when there is too little training data to train a BNN [57, 58]. With this new adaptation, we have to modify the update rule for the gradient descent, including the kernel

$$\theta^{t+1} = \theta^t + \alpha \nabla_{\theta^t} \log p(\theta^t|D) - \alpha \frac{\nabla_{\theta^t} \sum_{j=1}^n k(\theta^t, \theta_j^t)}{\sum_{i=1}^n k(\theta^t, \theta_i^t)} \,. \tag{6.21}$$

Throughout this thesis, we choose a Gaussian kernel written as

$$k(x, x') = \exp\left(-\frac{\sum_i (x_i - x_i')^2}{2}\right) \,. \tag{6.22}$$

Updating the ensemble members for every step during the training procedure can also be seen as a time evolution process, described using properties of ordinary differential equations (ODEs). Therefore, the extended update rule leads towards the ensemble of networks sampling the posterior distribution $\theta \sim p(\theta|D)$. To prove this posterior relation, we extend the update rule from Eq. (6.21) to a time-dependent probability density function $\rho(\theta, t)$. A similar procedure was done for the set up of conditional flow matching networks [282, 284] by describing the time evolution through an ODE or a continuity equation

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = v(\theta, t) \qquad \text{or} \qquad \frac{\partial \rho(\theta, t)}{\partial t} = -\nabla_\theta \left[v(\theta, t)\rho(\theta, t)\right] \,. \tag{6.23}$$

With a given velocity field $v(\theta, t)$, the individual paths $\theta(t)$ describe the evolving density $\rho(\theta, t)$. If the velocity field is chosen as

$$v(\theta, t) = -\nabla_\theta \log \frac{\rho(\theta, t)}{\pi(\theta)} \,, \tag{6.24}$$

the two equivalent conditions read

$$\begin{aligned} \frac{\mathrm{d}\theta}{\mathrm{d}t} &= -\nabla_\theta \log \frac{\rho(\theta, t)}{\pi(\theta)} \\ \frac{\partial \rho(\theta, t)}{\partial t} &= -\nabla_\theta \left[\rho(\theta, t) \nabla_\theta \log \pi(\theta)\right] + \nabla_\theta^2 \rho(\theta, t) \,. \end{aligned} \tag{6.25}$$

We can see that the continuity equations evolve into the Fokker-Planck equation, with $\rho(\theta, t) \to \pi(\theta)$ as the unique stationary probability distribution. A more detailed derivation for the repulsive ensemble can be found in Ref. [282].

As the next step, we can now relate the ODE from Eq. (6.23) to the previously determined update rule for repulsive ensembles, written down in Eq. (6.21). With this, the discretized version of the ODE becomes

$$\frac{\theta^{t+1} - \theta^t}{\alpha} = -\nabla_{\theta^t} \log \frac{\rho(\theta^t)}{\pi(\theta^t)} \,. \tag{6.26}$$

Since the explicit expression for the density $\rho(\theta^t)$ is unknown, it can be approximated by a superposition of kernels with correct normalization

$$\rho(\theta^t) \approx \frac{1}{n} \sum_{i=1}^n k(\theta^t, \theta_i^t) \quad \text{with} \quad \int \mathrm{d}\theta^t \rho(\theta^t) = \frac{1}{n} \sum_{i=1}^n \int d\theta^t k(\theta^t, \theta_i^t) = 1 \,. \tag{6.27}$$

Inserting this kernel approximation into the discretized ODE results in

$$\frac{\theta^{t+1} - \theta^t}{\alpha} = \nabla_{\theta^t} \log \pi(\theta^t) - \frac{\nabla_{\theta^t} \sum_i k(\theta^t, \theta_i^t)}{\sum_i k(\theta^t, \theta_i^t)} \,. \tag{6.28}$$

If we choose $\pi(\theta) \equiv p(\theta|D)$, we can identify the form of the discretized ODE with the update rule written in Eq. (6.21), confirming the correct convergence.

Until now, we encoded the repulsive ensemble only in weight space, but the desired predictions are provided in function space. For example, if two networks encode the same function by applying a permutation on the weights, they will not be affected by the repulsive force. Therefore, we have to introduce the repulsive term on the network output in the function space $f_\theta(x)$, translating it from weight space to function space by

$$\frac{f^{t+1} - f^t}{\alpha} = \nabla_{f^t} \log p(f|D) - \frac{\sum_j \nabla_{f^t} k(f, f_j)}{\sum_j k(f, f_j)} \,. \tag{6.29}$$

Nevertheless, the kernel can still be approximated as a Gaussian in a multidimensional space with an appropriately wide width. However, the network training is still defined in weight space, so the function space update rule has to be adapted to weight space using the appropriate Jacobian

$$\frac{\theta^{t+1} - \theta^t}{\alpha} = \nabla_{\theta^t} \log p(\theta^t|D) - \frac{\partial f^t}{\partial \theta^t} \frac{\sum_j \nabla_{f^t} k(f_{\theta^t}, f_{\theta_j^t})}{\sum_j k(f_{\theta^t}, f_{\theta_j^t})} \,. \tag{6.30}$$

Furthermore, we cannot evaluate the repulsive kernel in function space. Thus, we have to apply the same trick as for the BNN and evaluate the function for a finite batch of points x,

$$\frac{\theta^{t+1} - \theta^t}{\alpha} \approx \nabla_{\theta^t} \log p(\theta^t|D) - \frac{\sum_j \nabla_{\theta^t} k(f_{\theta^t}(x), f_{\theta_j^t}(x))}{\sum_j k(f_{\theta^t}(x), f_{\theta_j^t}(x))} \,, \tag{6.31}$$

As a last step, we have to translate the update rule from Eq. (6.31) into a loss function for the training of the RE. As for the BNN, we can obtain a tractable likelihood loss by applying Bayes' theorem, where we can neglect the evidence $p(D)$ but have to include the prior $p(\theta)$, which is again assumed to be Gaussian

$$\log p(\theta|D) = \log p(D|\theta) - \frac{\theta^2}{2\sigma^2} + \text{const.} \,, \tag{6.32}$$

where the second term represents the L2-normalization.

Given the training data set of size N, evaluated in batches of size B, f_{θ^t} is evaluated for all $x_1, ..., x_B$ samples in the batch. Also, not all occurrences of θ in the modified update rule are inside the gradient. To separate them clearly, we use the stop-gradient operation, denoted as $\hat{f}_{\theta_j}$. With these assumptions and modifications, we end up with the following loss function for a repulsive ensemble with n members

$$\mathcal{L}_{\text{RE}} = \sum_{i=1}^{n} \left[-\frac{1}{B} \sum_{b=1}^{B} \log p(x_b|\theta_i) + \frac{\beta}{N} \frac{\sum_{j=1}^{n} k(f_{\theta_i}(x), \hat{f}_{\theta_j}(x))}{\sum_{j=1}^{n} k(\hat{f}_{\theta_i}(x), \hat{f}_{\theta_j}(x))} + \frac{\theta_i^2}{2N\sigma^2} \right] \,. \tag{6.33}$$

The repulsive prefactor is usually chosen as $\beta = 1$. Again, a typical choice for the kernel is a Gaussian with a width given by the median heuristic [285]. With σ entering the loss

function in Eq. (6.33), REs are capable of estimating an uncertainty on the predictions. Thus, this additional parameter σ is connected to the repulsion. To provide us with a reliable uncertainty estimation of σ, the repulsion has to be applied in the log-likelihood space. This log-likelihood space is again used in the Amplitude regression task presented in Ch. 7.

6.3.3 Kolmogorov-Arnold networks

A KAN [286, 287] differs from a normal fully connected network in the sense that it does not learn weights but different activation functions for each layer. This is especially interesting for high-precision tasks because the activation function is responsible for adding non-linearities to the linear layers and thus affects the final prediction and the accuracy of the learned function. So far, only networks with static activation functions have been introduced, but KANs have the feature of learnable activation functions to better approximate the target distribution.

KANs are based on the Kolmogorov-Arnold representation theorem, which also underlies the deep set architecture [288] in a slightly different form. The theorem states that any multivariate smooth function with an n-dimensional input x and a function $f : [0,1]^n \to \mathbb{R}$, can be written as a finite composition of uni-variate functions and addition,

$$f(x) = f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) , \tag{6.34}$$

where $\phi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$. This theorem has downsides in practice since the uni-variate functions need to be non-smooth and/or fractal [289]. However, this problem can be resolved by expanding the decomposition into multiple layers, as shown in Ref. [286].

Consider a KAN consisting of L layers, with indices $l = 0, 1, ..., L-1$ and n_l dimensional input vector x_l to each layer. We then can define the action of each layer l via

$$x_{l+1,j} = \sum_{i=1}^{n_l} \phi_{l,j,i}(x_{l,i}) \qquad \text{with} \qquad j = 1, \ldots, n_{l+1} , \tag{6.35}$$

where the $n_{l+1} \times n_l$ dimensional functions $\phi_{l,j,i}$ are learnable. These functions could either be splines or rational functions [286, 290, 291].

Additionally, the sum in Eq. (6.35) can be turned into an operator matrix representation,

$$x_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\equiv \Phi_l} x_l , \tag{6.36}$$

where the function $\phi_{l,j,i}$ takes $x_{l,i}$ as input. Using this matrix representation, we can write down the full KAN network as

$$\mathrm{KAN}(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \ldots \circ \Phi_1 \circ \Phi_0) x , \tag{6.37}$$

with the form of a normal MLP network for comparison

$$\mathrm{MLP}(x) = (W_{L-1} \circ \mathrm{activation} \circ W_{L-2} \circ \ldots \circ \mathrm{activation} \circ W_0)x\,. \tag{6.38}$$

In this representation, the W_l are the linear operations, and the "activation" is a chosen fixed activation function. The use of KANs in physics is not entirely new; they have already been used in collider physics problems [292, 293].

As a middle ground between using full KAN networks or normal MLP networks, one can use GroupKAN networks [291]. They combine the features of simpler MLP networks with the advantage of learnable activation functions from the KAN architecture. Therefore, every layer gets replaced by a GroupKAN layer,

$$\mathrm{activation}(x) \to \mathrm{GroupKANlayer}_l(x) = \begin{pmatrix} \phi_{l,g_l(1)}(x_1) \\ \phi_{l,g_l(2)}(x_2) \\ \vdots \\ \phi_{l,g_l(n_l)}(x_{n_l}) \end{pmatrix}\,. \tag{6.39}$$

Notably, in the GroupKAN layer, sub-indices are introduced by grouping various functions together

$$g_l : \{1, 2, ..., n_l\} \to \{1, 2, .., m_l\}\,. \tag{6.40}$$

This can be read as $g_l(i) = k$ if i belongs to group k and m_l is the number of groups in the layer l. The number of groups can vary between one, only one common activation function is used for the full layer, and n, all entries have different activation functions.

This leads to the overall structure of the GroupKAN network with the form

$$\mathrm{GroupKAN}(x) = (W_{L-1} \circ \mathrm{GroupKANlayer}_{L-2} \circ W_{L-2} \circ \ldots \circ \mathrm{GroupKANlayer}_0 \circ W_0)x\,. \tag{6.41}$$

Comparing this GroupKAN structure to the structure of the full KAN network in Eq. (6.37), the GroupKAN implementation reduces the complexity significantly. Per layer in the KAN network, $n_{l+1} \times n_l$ functions have to be learned, while for the GroupKAN, only $m_l \leq n_l$ functions have to be learned. Still, the GroupKAN is more expressive than a normal MLP and can be inserted into any other architecture straightforwardly, making it far more versatile.

To conclude this chapter, we introduced the basic concepts of deep learning and the different types of uncertainties that need to be considered. We also introduced three different network architectures, namely BNNs, REs, and KANs. This knowledge is now put into action by applying it to the actual task of amplitude regression in Ch. 7, starting by using KANs as approach to learn the activation function and then testing the BNN's and RE's prediction precision and ability to learn systematic and statistical uncertainties.

Chapter 7

Accurate Uncertainty Estimation using Amplitude Regression

The research and the results presented in this chapter are based on work in collaboration with Henning Bahl, Luigi Favaro, Manuel Haußmann, Tilman Plehn, and Ramon Winterhalder and have been published in Ref. [3]. All figures and tables, as well as parts of the text, are similar or identical to the ones in the article.

We have introduced the basic concepts of a deep learning network and the loss function in Sec. 6.1. Based on these fundamental concepts, more advanced architectures like BNNs, REs, or KANs were discussed in Sec. 6.3. Next, these architectures are used on a physics regression task to test their accuracy in the prediction and the precision in estimating the different uncertainty types explained in Sec. 6.2.

The goal of particle physics is to identify the fundamental properties of particles within and beyond the SM. This goal can only be achieved by an optimal interplay of precision in predictions and measurements, which is constantly challenged by the vast amount of recorded and simulated data, its complexity, and its coverage of all aspects for optimal analysis. On the theory and simulation side of this challenge, we can use neural networks to improve and accelerate every aspect of the simulation chain, starting from phase-space sampling [39–42, 44, 294–297] to scattering amplitude evaluations [45–49, 51, 298–302], end-to-end event generation [284, 303–306], and ultra-fast detector simulations [307–328].

Mainly used networks for these tasks are surrogate and generative networks trained on simulated training data [329, 330]. Given the requirements from LHC data and performances, they have to function in a controlled and precise way over the entire phase space. [331–335]. With the LHC and the connected simulation chain generating larger amounts of data, the theory predictions shift to higher-order calculations, leading to the need to implement higher orders in perturbation theory. Here, we need to ensure that network surrogates trained on theory predictions are controlled and precise enough to reflect the precision of the underlying theory. Surrogates or density estimators can learn uncertainties on the network prediction, for instance, as BNNs [54, 336, 337], REs [56–58], or likelihood methods [338, 339]. Even if the learned uncertainty is not used in a downstream analysis, a reliable uncertainty estimate is key to the justification for using an ML surrogate.

Therefore, we use surrogate amplitudes encoded in modern neural networks [45–49, 51, 298–302] to evaluate higher-order scattering amplitudes as fast as tree-level amplitudes. With this aspect, the challenge for these networks is that perturbative quantum field theory requires them to reliably reproduce the theory prediction on the training data

with sufficient accuracy. We study limitations to precision surrogates, the accuracy, and the control via a learned uncertainty.

For that reason, we already introduced our different architectures in Sec. 6.3, namely a deterministic network using a heteroscedastic loss function to estimate systematic uncertainties, a BNN to learn statistical and systematic uncertainties, and an RE to learn statistical uncertainties. Statistical uncertainties are defined as vanishing for infinite amounts of training data, while systematics can have many sources related to the training data, the network architecture, or the network training. The latter are the limiting factor in LHC applications. Thus, we are highly interested in learning systematic uncertainties, assuming that statistical limitations are negligible.

In Sec. 7.1, we adapt the different architectures, already generally introduced in Sec. 6.3, to the amplitude regression task and the concrete definition of these different uncertainties for the networks. Next, Sec. 7.2 introduces our dataset of loop-induced amplitudes for the partonic process $gg \to \gamma\gamma g$ [47, 49], and in Sec. 7.3, we perform a systematic study of the effect of activation functions on the accuracy of amplitude learning, using KANs. This is followed by a deeper look into uncertainties induced by adding artificial noise to our dataset in Sec. 7.4 and the calibration of the learned statistical uncertainties in Sec. 7.5.

7.1 Learned uncertainties

In this project, we train a regression network to learn the transition amplitude or matrix element squared $A(x)$ as a function of phase space x,

$$A_{\text{NN}}(x) \approx A_{\text{true}}(x) \, . \tag{7.1}$$

The key property of a transition amplitude is that A_{true} is exact at a given order in perturbation theory, i.e. there are no stochastic processes entering the prediction, making it ideal for an uncertainty and precision study. The training data D_{train} consists of phase space points and their corresponding amplitude values $(x, A)_j$.

7.1.1 Heteroscedastic loss

We now denote the network parameters as θ, and the network training is maximizing the probability of the network parameters to describe the training data, $p(\theta|D_{\text{train}})$. Because we do not have access to this probability, we use Bayes theorem and instead minimize the negative log-likelihood on a set of amplitudes

$$\mathcal{L} = -\Big\langle \log p(A|x,\theta) \Big\rangle_{x\sim D_{\text{train}}} \, . \tag{7.2}$$

The form of the log-likelihood depends on the nature of the training data. As shown in Sec. 6.1 for including systematic uncertainties, the natural choice is a Gaussian with heteroscedastic variance. Thus, we can specify the general heteroscedastic loss from Eq. (6.10) to the amplitude dataset with $A_{\text{NN}}(x)$ and $\sigma(x)$ being the two functions over the phase space

$$\mathcal{L}_{\text{het}} = \left\langle \frac{|A_{\text{true}}(x) - A_{\text{NN}}(x)|^2}{2\sigma(x)^2} + \log \sigma(x) \right\rangle_{x\sim D_{\text{train}}} \, . \tag{7.3}$$

As mentioned, the variance $\sigma^2(x)$ is unknown and must be learned. Therefore, we have to include a normalization to drive $\sigma(x)$ towards large values. For this work, we have tested that it is sufficient to assume a Gaussian likelihood, but the heteroscedastic loss can be extended to a Gaussian mixture model for $\sigma(x)$ [336, 340].

Even without using the built-in uncertainty, it has the advantage of acting as a cutoff, allowing the network to ignore unlearnable aspects and features to focus on other features without overtraining.

However, a heteroscedastic loss does not capture uncertainties arising from the actual training. This is why we also include a BNN and an RE for quantifying per-amplitude uncertainty.

7.1.2 Bayesian neural network

A full explanation and derivation of the loss function for a BNN is provided in Sec. 6.3.1. Here, we mainly focus on the motivation for using BNNs in regression tasks, adapting the loss function in Eq. (6.15) to this specific task, and the derivation of uncertainties.

Regarding LHC tasks, BNNs can be used e.g., for amplitude regression [49], jet calibration [337], classification [336], and generative networks [284, 331, 341, 342].

Assuming that the amplitude for a given phase space point is following a probability distribution $p(A|x)$ with mean

$$\langle A \rangle(x) = \int \mathrm{d}A \; A \; p(A|x) \; , \tag{7.4}$$

the network will encode this probability as weight configurations, conditional on the training data. Following the derivation in Sec. 6.3.1, we get the following expression for the KL-divergence

$$\begin{aligned} \mathrm{KL}[q(\theta), p(\theta|D_\text{train})] = &\mathrm{KL}[q(\theta), p(\theta)] - \int \mathrm{d}\theta \; q(\theta) \log p(D_\text{train}|\theta) \\ &+ \log p(D_\text{train}) \int \mathrm{d}\theta \; q(\theta) \; . \end{aligned} \tag{7.5}$$

Again, the first term acts as regularization, while the second one represents the negative log-likelihood. The third term simplifies to $\log p(D_\text{train})$, the evidence of D_train. It is a constant with respect to θ, so the modified BNN loss compared to Eq. (6.15) is

$$\mathcal{L}_\text{BNN} = \mathrm{KL}[q(\theta), p(\theta)] - \Big\langle \log p(D_\text{train}|\theta) \Big\rangle_{\theta \sim q(\theta)} \; . \tag{7.6}$$

We are free to choose the prior $p(\theta)$, so we follow common practice and choose independent Gaussians with a given width for each weight.

Uncertainties

To extract the uncertainty for $A(x)$, we rewrite Eq.(7.4) such that we sample over θ and define an expectation value and the corresponding variance

$$\langle A \rangle(x) = \int \mathrm{d}\theta \; q(\theta) \; \overline{A}(x, \theta) \qquad \text{with} \qquad \overline{A}(x, \theta) = \int \mathrm{d}A \; A \; p(A|x, \theta)$$

$$\begin{aligned}\sigma^2_{\text{tot}}(x) &= \int \mathrm{d}\theta\; q(\theta) \left[\overline{A^2}(x,\theta) - \overline{A}(x,\theta)^2 + \Big(\overline{A}(x,\theta) - \langle A\rangle(x)\Big)^2\right] \\ &\equiv \sigma^2_{\text{syst}}(x) + \sigma^2_{\text{stat}}(x)\;, \end{aligned} \tag{7.7}$$

where $\overline{A^2}(x,\theta)$ is defined in analogy to $\overline{A}(x,\theta)$. Thus, the total uncertainty can be factorized into two terms. The first,

$$\sigma^2_{\text{syst}}(x) \equiv \int \mathrm{d}\theta\; q(\theta)\; \sigma(x,\theta)^2 = \int \mathrm{d}\theta\; q(\theta)\; \left[\overline{A^2}(x,\theta) - \overline{A}(x,\theta)^2\right]\;,$$

corresponds to the learned error in the heteroscedastic loss in Eq.(7.3). Given exact $A_{\text{true}}(x)$, it vanishes in the limit of arbitrarily well-known data and perfect network training

$$p(A|x,\theta) = \delta(A - A_{\text{true}}(x)) \qquad \Leftrightarrow \qquad \overline{A^2}(x,\theta) = A_{\text{true}}(x)^2 = \overline{A}(x,\theta)^2\;. \tag{7.8}$$

We will see that it approaches a plateau for large training datasets, so we refer to it as a systematic uncertainty—accounting for noisy data or labels [337], limited expressivity of the network (structure uncertainty) [49], non-optimal network architectures in the presence of symmetries, non-smart choices of hyperparameters or any other sources of systematic uncertainty.

The second error is the θ-sampled variance

$$\sigma^2_{\text{stat}}(x) = \int \mathrm{d}\theta\; q(\theta)\; \left[\overline{A}(x,\theta) - \langle A\rangle(x)\right]^2\;, \tag{7.9}$$

It vanishes in the limit of perfect training, leading to uniquely defined network weights θ_0,

$$q(\theta) = \delta(\theta - \theta_0) \qquad \Leftrightarrow \qquad \langle A\rangle(x) = \overline{A}(x,\theta_0)\;, \tag{7.10}$$

and is thus called statistical uncertainty.

For small training datasets, these two uncertainties cannot be easily separated, but we can separate them clearly for sufficiently large training datasets [336], where σ_{stat} approaches zero, while the systematic error σ_{syst} reaches a finite plateau. Usually, in LHC physics, we can make sure to use enough training data such that

$$\sigma_{\text{tot}}(x) \approx \sigma_{\text{syst}}(x) \gg \sigma_{\text{stat}}(x)\;. \tag{7.11}$$

Correspondingly, we focus on the extraction and validation of learned systematic uncertainties.

7.1.3 Repulsive ensembles

An alternative way to compute the uncertainty on a network output is using (repulsive) ensembles. An in-detailed derivation and explanation for these can be found in Sec. 6.3.2 of this thesis. Thus, they will only be briefly discussed in the scope of uncertainty estimation and their adaption to the amplitude regression tasks.

Adapting the general loss function for REs in Eq. (6.33) to this specific task using the

Amplitudes $A(x)$, the new loss function for n repulsive ensembles read as

$$\mathcal{L}_{\text{RE}} = \sum_{i=1}^{n} \left[-\frac{1}{B} \sum_{b=1}^{B} \log p(A|x_b, \theta_i) + \frac{\beta}{N} \frac{\sum_{j=1}^{n} k(A_{\theta_i}(x), \hat{A}_{\theta_j}(x))}{\sum_{j=1}^{n} k(\hat{A}_{\theta_i}(x), \hat{A}_{\theta_j}(x))} + \frac{\theta_i^2}{2N\sigma^2} \right] . \tag{7.12}$$

For a training dataset of size N, evaluated for B batches, $A_{\theta_i}(x)$ is evaluated for all $x_1, ..., x_B$. Again, not all occurrences of θ are inside the gradient, so we use the stop-gradient operation, denoted as $\hat{A}_{\theta_j}(x)$. The kernel is chosen as a Gaussian with a width given by the median heuristic [285]. The network outputs not only the predicted amplitude but also the learned variance. We have to select in which space the repulsion is applied. In this application, we use the log-likelihood space. As a note, applying the repulsion to the amplitudes and variances separately does not lead to different results.

7.1.4 Calibration and pulls

To determine if the learned uncertainty is correctly calibrated, we can look at pull distributions. Let us start with a deterministic network learning $A_{\text{NN}}(x) \approx A_{\text{true}}(x)$ using a heteroscedastic loss. Given the learned local uncertainty $\sigma(x)$ by the network, we can evaluate the pull of the learned combination as

$$t_{\text{het}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{\sigma(x)} , \tag{7.13}$$

If $\sigma(x)$ captures the absolute value of the deviation of the learned function from the truth exactly and for any x-value, then the pull would be

$$A_{\text{NN}}(x) = A_{\text{true}}(x) \pm \sigma(x) \qquad \Leftrightarrow \qquad t(x) = \pm 1 . \tag{7.14}$$

Realistically, an uncertainty estimate will only capture the maximum deviation, so the per-amplitude deviation of $|A_{\text{NN}}(x) - A_{\text{true}}(x)|$ might be smaller. For a stochastic source of uncertainties, the learned values $A_{\text{NN}}(x)$ are assumed to follow a Gaussian distribution around the mean $\langle A_{\text{NN}}(x) \rangle$. The width of this Gaussian should be given by the learned uncertainty, which means the pull will follow a unit Gaussian.

As an example, let us assume that we learn $A_{\text{true}}(x)$ from noisy training data,

$$A_{\text{true}}(x) \;\to\; A_{\text{train}}(x) \qquad \text{with} \qquad p(A_{\text{train}}(x)) = \mathcal{N}(A_{\text{true}}(x), \sigma^2_{\text{train}}(x)) . \tag{7.15}$$

We then minimize the heteroscedastic log-likelihood loss from Eq.(7.3) based on this smeared input amplitudes $A_{\text{train}}(x)$. If the network were allowed to overfit perfectly, the outcome would be $A_{\text{NN}}(x) = A_{\text{train}}(x)$. In that case, $\sigma(x)$ is not needed for the loss minimization and hence not learned at all. If we keep the network from overtraining, the best that the network can learn from some unbiased data is

$$A_{\text{NN}}(x) \approx A_{\text{true}}(x) \qquad \text{and} \qquad \sigma(x) \approx \sigma_{\text{train}}(x) . \tag{7.16}$$

In that case, we can define a pull function over phase space as

$$t_{\text{het}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{train}}(x)}{\sigma(x)} . \tag{7.17}$$

When the input noise is learned as the network uncertainty, it follows a unit Gaussian.

Note that the pull compares the learned function with the training data, not with the idealized data in the limit of zero noise.

Systematic pull

We can translate this result for a deterministic network with a heteroscedastic loss to the BNN and variational inference in the limit $q(\theta) = \delta(\theta - \theta_0)$,

$$\langle A \rangle(x) = \int \mathrm{d}\theta dA \; A \; p(A|x,\theta) \; q(\theta) = \int \mathrm{d}A \; A \; p(A|x,\theta_0) \; . \tag{7.18}$$

The network is trained well if

$$p(A|x,\theta_0) \approx p(A|x) \; . \tag{7.19}$$

In the Gaussian case, this is equivalent to learning its mean and width

$$\langle A \rangle(x) \approx A_\text{true}(x) \qquad \text{and} \qquad \sigma_\text{syst}(x) \approx \sigma_\text{train}(x) \; . \tag{7.20}$$

As argued above, the systematic pull relates the learned $A(x)$ to the actual and potentially noisy training data $A_\text{train}(x)$,

$$\begin{aligned} t_\text{syst}(x) &= \frac{1}{\sigma_\text{syst}(x)} \int dA \; \left[A - A_\text{train}(x)\right] \; p(A|x,\theta_0) \\ &= \frac{1}{\sigma_\text{syst}(x)} \left[\int \mathrm{d}A \; A \; p(A|x,\theta_0) - A_\text{train}(x) \int \mathrm{d}A \; p(A|x,\theta_0) \right] \\ &= \frac{\langle A \rangle(x) - A_\text{train}(x)}{\sigma_\text{syst}(x)} \; . \end{aligned} \tag{7.21}$$

Notably, we do not have to approximate the integral in the definition of $t_\text{syst}(x)$ since the network directly predicts the parameters of $p(A|x,\theta_0)$.

Statistical pull

To define a pull, equivalently to the one of the systematic uncertainty, to test the calibration of the statistical uncertainty, we remind ourselves that we need to extract the statistical uncertainties defined in Eq.(7.9) using the unbiased sample variance from N amplitudes sampled from the posterior weight distributions, $\theta_i \sim q(\theta)$,

$$\begin{aligned} \langle A \rangle(x) &\approx \frac{1}{N} \sum_i \; \overline{A}(x,\theta_i) \\ \sigma^2_\text{stat}(x) &\approx \frac{1}{N-1} \sum_i \; \left[\overline{A}(x,\theta_i) - \langle A \rangle(x) \right]^2 \; . \end{aligned} \tag{7.22}$$

This definition provides us with the pull distribution

$$\hat{t}_\text{stat}(x,\theta) = \frac{\overline{A}(x,\theta) - \langle A \rangle(x)}{\sigma_\text{stat}(x)} \; . \tag{7.23}$$

It samples amplitudes from the posterior, and it calculates the deviation from the expectation value $\langle A \rangle$ and the uncertainty $\sigma_\text{stat}(x)$, both computed by sampling θ. This

pull is guaranteed to follow a standard Gaussian for large N, as $\sigma_{\text{stat}}(x)$ is calculated from the variance of $\overline{A}(x,\theta)$. Under the assumption of perfect training, we can replace the expectation value with its learning target, $\langle A\rangle(x) \approx A_{\text{true}}(x)$. To take all the different ensemble members into account, we define the pull

$$t_{\text{stat}}(x,\theta) = \frac{\overline{A}(x,\theta) - A_{\text{true}}(x)}{\sigma_{\text{stat}}(x)} \,. \tag{7.24}$$

This pull should again be a standard Gaussian, which means we can use it to test the calibration of the learned statistical uncertainty $\sigma_{\text{stat}}(x)$.

Scaled pull

The problem with Eq.(7.24) is that the pull depends on θ, which means it can only be computed for individual members of the BNN or RE, but not for the sampled set of amplitudes. For a global pull, we would need to replace $\overline{A}(x,\theta)$ with a prediction after sampling and averaging over θ.

Starting with systematics, we can already define the underlying problem when we extract the prediction and the uncertainty from ensembles. When we evaluate the ensemble members together, ensembling improves the central value, but the learned uncertainty does not benefit from the ensemble structure, which will become clearer in the following sections. In the next section, we will indeed see that ensembling, without or with a repulsive kernel, will lead to poorly calibrated, under-confident systematic pulls.

Moving on to statistical uncertainties, when extracting the expectation value and the variance from M samples, we know the scaling. Assuming identically distributed variables with unit weights, taking means for the prediction should reduce the statistical error by a factor $\sqrt{M}$, so we define the scaled standard deviation as

$$\sigma_{\text{stat},M}(x) = \frac{\sigma_{\text{stat}}(x)}{\sqrt{M}} \,. \tag{7.25}$$

In analogy to Eq.(7.23), we can define the corresponding averaged pull as

$$\hat{t}_{\text{stat},M}(x) = \frac{\langle A\rangle_M(x) - \langle A\rangle(x)}{\sigma_{\text{stat},M}(x)} \,. \tag{7.26}$$

To test the calibration, we would again replace the full expectation value with its learning target, $\langle A\rangle(x) \approx A_{\text{true}}(x)$ and define

$$t_{\text{stat},M}(x) = \frac{\langle A\rangle_M(x) - A_{\text{true}}(x)}{\sigma_{\text{stat},M}(x)} \,. \tag{7.27}$$

Under the above conditions, it should follow a standard Gaussian. By varying M, we can interpolate between the member-wise pull of Eq.(7.24) for $M = 1$ and the fully sampled pull with maximal M. We will see later that it starts deviating significantly beyond the single-element case $M = 1$. This does not imply that the statistical uncertainties are poorly calibrated but that the two samplings are not (sufficiently) independent.

7.2 Amplitude data and network architectures

As the benchmark for our surrogate amplitudes we choose the squared loop-induced matrix elements for the partonic process [47, 49]

$$gg \to \gamma\gamma g \ , \tag{7.28}$$

where we generate unweighted events with SHERPA [343], and then obtain the corresponding amplitudes with the NJET library [344]. A set of basic cuts on the partons,

$$\begin{aligned} &p_{T,j} > 20 \text{ GeV} && |\eta_j| < 5 && R_{jj,j\gamma,\gamma\gamma} > 0.4 \\ &p_{T,\gamma} > 40, 30 \text{ GeV} && |\eta_\gamma| < 2.37 \ , \end{aligned} \tag{7.29}$$

mimics the detector acceptance and object definition. The total size of the dataset is 1.1M phase space points with their corresponding amplitudes. Unless explicitly mentioned, we only use 70% of the data for training and train for 1000 epochs.

The accuracy of the network prediction is measured locally by

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)} \ . \tag{7.30}$$

To illustrate the accuracy of the networks, we histogram these values for a test dataset containing 20% of the complete dataset. The remaining 10% is used for validation and the selection of the best network. The width of the histogrammed Δ-values for the test dataset provides the accuracy of the surrogate amplitude.

Neural networks

The following sections include detailed studies on the accuracy and learned uncertainties of several network architectures. With increasing complexity, we use the following network architectures:

- an MLP, a fully connected network with linear layers followed by non-linearities, also labeled as Det;
- an MLP-I, like the MLP but including Mandelstam invariants as additional input features, also labeled as Det-I;
- a Deep Sets (DS) network [288], which learns an embedding for each particle type;
- a Deep Sets Invariants (DSI) network, i.e. DS with Mandelstam invariants as additional inputs;
- an L-GATr network, a fully Lorentz and permutation-equivariant network architecture [299, 301].

In the following, the MLP networks are used for the BNN and RE results. The DS network introduces an embedding step before a standard MLP. It is shared across particles of the same type, and it is implemented as a fully connected network with a final representation vector of dimension 64. All the representations are concatenated before being passed to the second fully connected network, which predicts the amplitudes. The DSI version uses a structure similar to the DS, where the input four-vectors are concatenated to Lorentz invariants. It is a specialized architecture for LHC amplitudes [299], combining a deep sets

architecture [288] — providing permutation invariance — with Lorentz invariants [299]. As an additional test, we compared these architectures also to an MLP-I network, where we explored the effect of only using Lorentz invariants to train the MLP network

As mentioned above, for both the MLP-I and the DSI, we augment all possible combinations of Mandelstam invariants from the input four-vectors,

$$s_{ij} = (p_i + p_j)^2 = 2p_i p_j \ , \tag{7.31}$$

which we additionally transform with a logarithm to obtain $\mathcal{O}(1)$ quantities that are easier to handle by neural networks. The L-GATr architecture can be used for amplitude regression [51, 58, 299, 301]. We use a modified version of the original network, adding the heteroscedastic loss and the learned systematic uncertainties. The choice of hyperparameters for all the networks — if not stated explicitly in the text — can be found in App. C.

7.3 KAN amplitudes

As previously discussed in Sec. 6.3.3 KANs [286, 287] are an ideal representation to systematically test different types of non-linearities. Thus, we perform a systematic study of the effect of different types of non-linearities on fixed non-linear activation functions as part of the network training.

In addition to the KAN network, structured as

$$\mathrm{KAN}(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \ldots \circ \Phi_1 \circ \Phi_0)x \ , \tag{7.32}$$

we also test GroupKAN network architectures, introduced in Ref. [291]. These are in between a fully implemented KAN architecture and normal MLP networks. Thus, they can be viewed as normal MLP networks with learnable activation functions in between, making it easier to test and train different implementations and types of non-linearities. The overall structure of a GroupKAN network can be written as

$$\mathrm{GroupKAN}(x) = (W_{L-1} \circ \mathrm{GroupKANlayer}_{L-2} \circ W_{L-2} \circ \ldots \circ \mathrm{GroupKANlayer}_0 \circ W_0)x \ , \tag{7.33}$$

which reduces the overall complexity and training time compared to a KAN.

7.3.1 Activation functions

To test the effect of various activation functions not only on a single architecture, we investigate and compare three different architectures. The first one is a simple MLP with three hidden layers and 128 hidden dimensions. For this MLP architecture, we compare the use of different non-linearities, three fixed activation functions, such as ReLU, GELU, and leaky ReLU, to a GroupKAN approach with 1, 2, 4, or 8 groups, which we denote as "GroupKAN-1" etc. We parameterize the learnable GroupKAN activation functions using rational functions [290, 291] with an order five polynomial in the numerator and denominator, respectively. The used MLP networks have $\sim 4 \cdot 10^4$ parameters. Additionally, we also test a DSI architecture with $\sim 2 \cdot 10^5$ parameters by again comparing several fixed activation functions to the GroupKAN approach. Moreover,

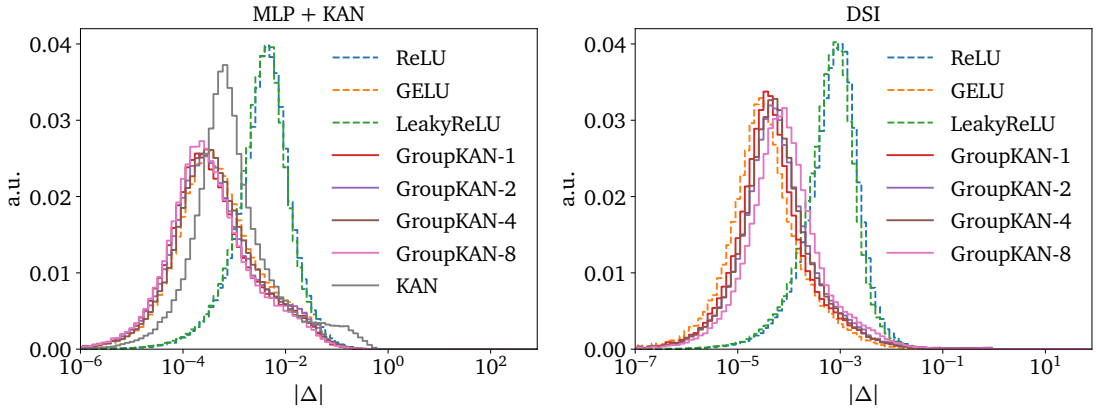

Figure 7.1: Accuracy on a logarithmic scale for the MLP, DSI, and KAN networks.

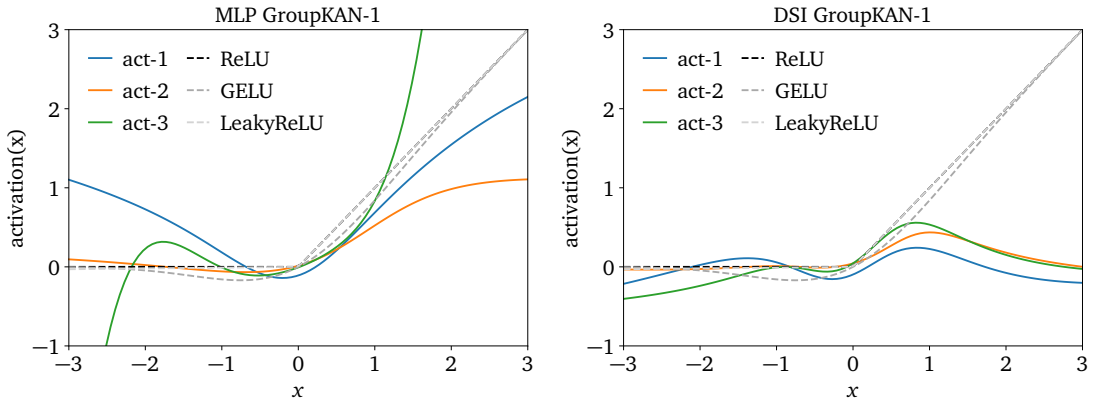

Figure 7.2: Learned activation functions for the GroupKAN-1 MLP and DSI networks.

we test a full KAN network with three hidden layers and 64 hidden dimensions. We use B-splines with a polynomial order of three for the learnable functions and a grid size of ten. The KAN network has $\sim 1 \cdot 10^5$ parameters.

As a result, we show $|\Delta|$ distributions for the various MLP and KAN networks in the upper left panel of Fig. 7.1. There, we can see that the GroupKan networks provide us with the most accurate predictions, with the bulk of the distribution lying between 10^{-4} and 10^{-3}. We observe no significant performance increase when the number of groups is increased. The GELU network is slightly worse than the GroupKAN network. In contrast, the ReLU and LeakyReLU networks have significantly worse performance with $|\Delta|$ values centered around $\sim 10^{-2}$. The KAN network lies in between the GroupKAN and ReLU/LeakyReLU results. Notably, the KAN distribution has a shoulder for large $|\Delta|$, indicating that the prediction is significantly worse for a subset of all amplitudes.

Next, we turn to the DSI architecture as shown in the right panels of Fig. 7.1. We again observe that the GroupKAN and the GELU networks have similar performance, while the ReLU and LeakyReLU networks perform significantly worse. In contrast to the MLP networks, the network using a GELU as an activation function, however, slightly outperforms the GroupKAN networks.

We provide a visualization of the learned activation functions for the GroupKAN-1 networks in Fig. 7.2. In the left panel, the MLP GroupKAN-1 activation functions

for the three layers are shown in comparison to the tested fixed activation functions. The learned activation functions behave similarly to the fixed activation functions for $x \sim 0$ but deviate significantly for larger $|x|$ values. The behavior is similar for the DSI GroupKAN-1 network, for which we show the activation layer of the summary net in the right panel of Fig. 7.2. Notably, the learned activation functions are flatter than for the MLP case.

In conclusion, we find full KAN networks to provide no significant improvement over architectures with fixed activation functions. GroupKAN layers can, however, be a useful tool to enhance the accuracy of MLP networks. For more complex architectures like the DSI architecture, we find no improvement above a well-chosen fixed activation function. They, nevertheless, can be helpful in establishing a basis line if an extensive scan of different fixed activation functions is not feasible.

From the MLP and DSI GroupKAN study, Fig. 7.2 shows that there is a slight preference towards choosing the GELU as an activation function compared to the ReLU. Therefore, we compared the impact of both activation functions on a normal MLP network labeled as Det. Fig 7.3 shows this comparison, indicating a clearly more precise prediction when using the GELU activation function compared to the ReLU function.

7.4 Systematics

From the two sources of uncertainties, statistics and systematics, we start by studying the systematics. They can be extracted through the heteroscedastic loss, which can be included in the BNN and the RE. For LHC applications, like amplitude regression, this uncertainty type dominates the total uncertainty. In this section, we study the surrogate limitations traced by systematic limitations. Therefore, we start by introducing artificial noise and then move on to study the network expressivity and symmetry-aware network architectures.

Unless mentioned explicitly, we use a BNN prior variance $\sigma_{\text{prior}} = 1$ and a repulsive prefactor of $\beta = 1$ for the repulsive ensemble.

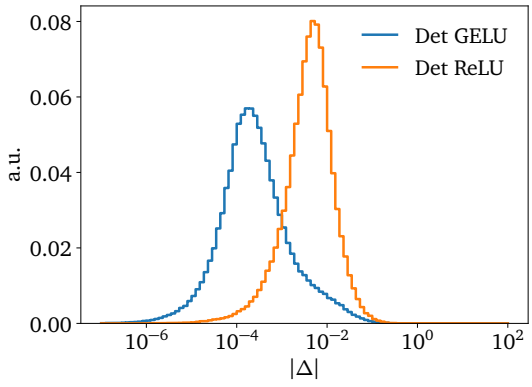

Figure 7.3: Comparing GELU and ReLU as possible activation functions for a deterministic network (Det).

7.4.1 Systematics from noise

There are many sources of systematic uncertainties, and as a starting point, we apply Gaussian noise to the data set to determine if the noise can be learned by the networks [337]. Using Eq.(7.15), we define the artificial noise level relative to the amplitude,

$$A_\text{train} \sim \mathcal{N}(A_\text{true}, \sigma^2_\text{train}) \qquad \text{with} \qquad \sigma_\text{train} = f_\text{smear}\, A_\text{true}\ , \tag{7.34}$$

where we consider the relative noise fractions

$$f_\text{smear} = \{0.25, 0.5, 0.75, 1, 2, 3, 5, 7, 10\}\% \ . \tag{7.35}$$

This introduces stochastic systematics in the data. Assuming that the additional systematics factorizes, it should appear as added to the other uncertainties in quadrature,

$$\sigma^2_\text{tot} = \sigma^2_\text{syst} + \sigma^2_\text{stat} = \sigma^2_\text{syst,0} + \sigma^2_\text{noise} + \sigma^2_\text{stat} \qquad \Leftrightarrow \quad \sigma^2_\text{noise} = \sigma^2_\text{syst} - \sigma^2_\text{syst,0}\ . \tag{7.36}$$

The contribution $\sigma_\text{syst,0}$ represents the systematic uncertainty stemming, for instance, from a limited network expressivity, where σ_noise indicates the artificially added uncertainty from the noise.

In Fig. 7.4 we show the two learned uncertainties from the BNN as a function of the amount of training data. Each point is the median of the respective uncertainty extracted from a test sample of 200K amplitudes. We perform this scan for the noise levels $\sigma_\text{smear} = \{5, 2, 0\}\%$ (left to right). For the largest noise rate, we see that the learned statistical uncertainty vanishes towards the large training sample, leaving us with a well-defined plateau for the systematic uncertainty. On the other hand, we also see that the split into statistical and stable systematic uncertainties only applies in the limit of large training samples and vanishing σ_stat. For finite training data size, the systematic uncertainty depends on the training sample size as well. The reason is that, for limited training data, systematic uncertainties can occur when the network's, in principle, sufficient expressivity is not fully explored after limited training time.

When we reduce the artificial noise to 2%, the learned systematics plateau drops to the corresponding values, and without added noise, it reaches a finite plateau below 1% relative systematics. It represents the next level of systematic uncertainty. We will see in Sec. 7.4.2 that it is related to the expressivity or size of the regression network.

In Fig. 7.5, we confirm that the learned systematics indeed reproduce the artificial noise added to the training data. In the left panel, we show the extracted uncertainties for

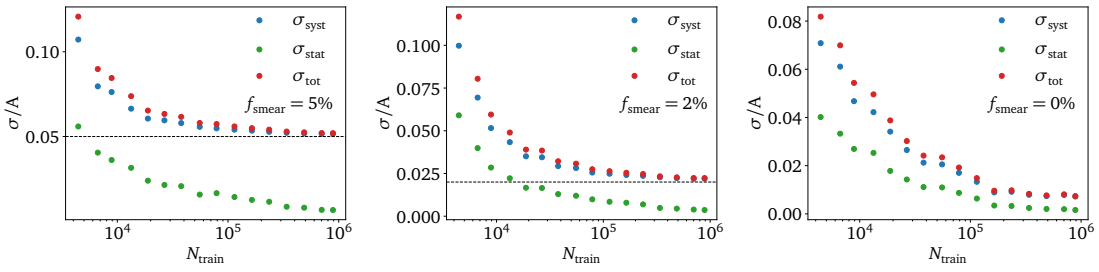

Figure 7.4: Relative systematic and statistical uncertainties learned by the BNN as a function of the dataset size, for 5%, 2%, and zero artificial noise on the training data.

the BNN, the REs, and a deterministic network with a heteroscedastic loss. All three methods agree for the systematic type of uncertainty, including the feature that for noise below 2%, the learned systematic uncertainty approach a new target value around 0.5% of relative systematic uncertainty.

For the BNN and the RE, we also look at the learned statistical uncertainty, which should be independent of the added noise. For the BNN, the learned statistical uncertainty is below 0.1% in the no-noise limit. Adding noise makes this estimate less reliable, which reflects the numerical problem of separating two contributions and adding in quadrature if one of them is a factor 100 larger than the other. Interestingly, the statistical uncertainty learned by the REs is significantly larger. Because the data efficiency of the two network training can, in principle, be different, this is expected, and we will discuss the learned statistical uncertainties more in Sec. 7.5

In the right panel of Fig. 7.5, we show the calibration curve for the input vs learned noise using the definition of Eq.(7.36). It confirms the excellent and consistent behavior of the three different implementations. The differences in the learned statistical uncertainties are numerically too small to affect the calibration significantly. As a word of caution — in the Appendix, we also show the extracted noise when we apply REs without a heteroscedastic loss. While one might speculate that in this setup, the REs would still learn the systematic uncertainty, it really does not. REs without heteroscedastic loss are really limited to statistical uncertainties.

The correlation shown in Fig. 7.5 indicates that the learned systematics extract the added noise correctly. However, the correlation only shows the median uncertainty, so we still want to check the uncertainty distributions using the systematic pulls introduced in Eq.(7.21). In Fig. 7.6, we show the relative accuracy and the systematic pull distributions for three different noise levels. If stochastic systematics are learned correctly, the pull should follow a unit Gaussian. In the upper panels, we first confirm that the accuracy improves with less noise, first consistently for the three methods, and then towards zero noise with different accuracies for the REs on the one side, the BNN, and the heteroscedastic noise. In all cases, the relative accuracy distributions follow an approximate Gaussian curve.

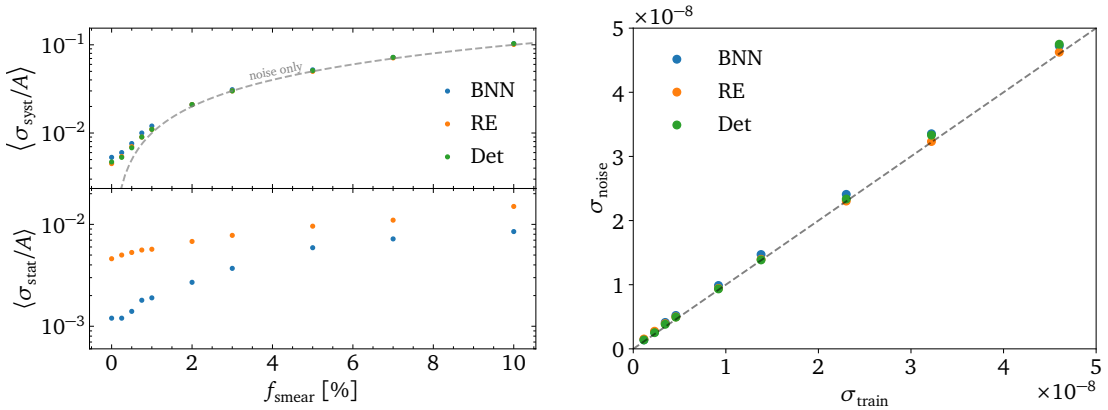

Figure 7.5: Left: Relative uncertainty versus artificial noise for different network architectures. The "noise only" curve shows the scaling of the systematics, assuming the added noise is the only source of uncertainty. The exact numbers are given in Tab. C.2. Right: Extracted noise, defined in Eq.(7.36), as a function of the input noise for BNN, REs, and a heteroscedastic deterministic network.

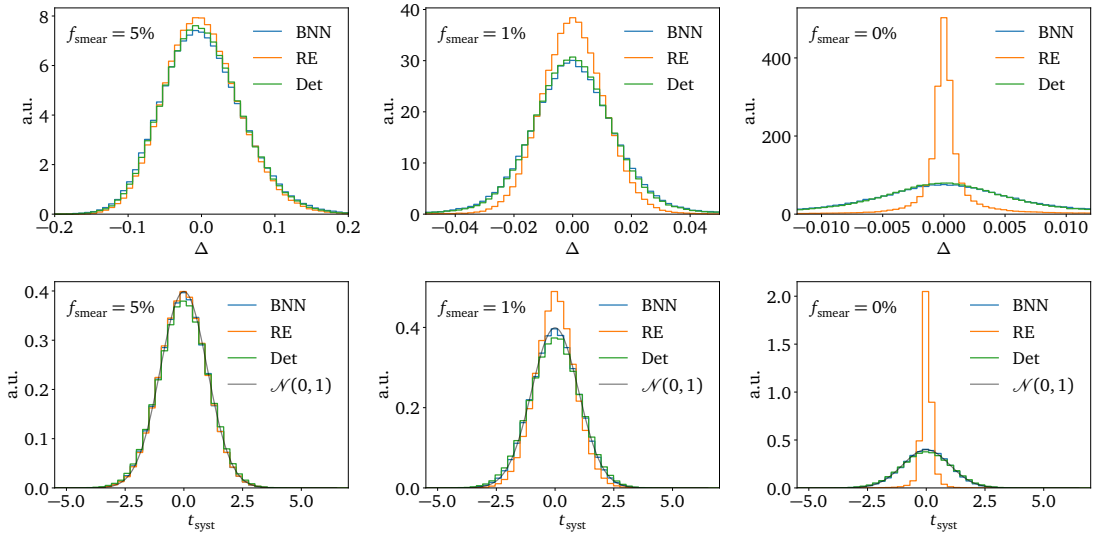

Figure 7.6: Relative accuracy (upper) and systematic pull (lower) for the BNN, REs, and heteroscedastic loss for decreasing added noise.

In the second row of Fig. 7.6, we show the systematic pull, combining the accuracy in the numerator with the learned uncertainty in the denominator. This combination should become a universal unit Gaussian. The BNN and the heteroscedastic network indeed reproduce this pattern for all added noise levels. In the zero noise limit, the RE gives a too-narrow systematic pull distribution, indicating that the learned systematics from the heteroscedastic modification of the REs are too conservative. From Fig. 7.5, we can speculate that the too-large learned systematics in the limit of zero noise is related to the fact that the REs extract a common $\sigma_{\text{syst}} \approx \sigma_{\text{stat}} \approx 0.5\%$ in this limit. The heteroscedastic loss in the REs only extracts the correct systematics when it is larger than the learned statistical uncertainty.

Calibrating ensembles

In Fig. 7.6, we observe a poor calibration of (repulsive) ensembles. We have checked that the repulsive kernel has hardly any impact on our setup, so we can look at standard ensembles to understand this issue. The left panel of Fig. 7.7 confirms that the different initializations of the ensemble members lead to convergence in different local minima, which may predict some amplitudes better than others. Extracted independently for each member, and with a heteroscedastic loss, we see that the pull is perfectly calibrated.

In Sec. 7.1.4, we have already discussed that ensemble training improves the accuracy of the amplitude regression, but the learned uncertainties do not benefit from them. This is confirmed by the right panel of Fig. 7.7, where the RE pull from a larger ensemble becomes increasingly too narrow. While the mean network prediction becomes more accurate, the learned σ_{syst} only accounts for the systematics of the single network, which leads to the poorly calibrated systematic RE pull.

Because the narrow pull distributions as Gaussian do not show any bias, we can solve this issue using a targeted calibration step. The simplest solution introduces a single global parameter, which rescales all the amplitudes and effectively changes the parameters of each learned Gaussian as $\sigma_{\text{syst}} \to \sigma_{\text{syst}} \times T$. Fig. 7.8 shows the systematic pull before

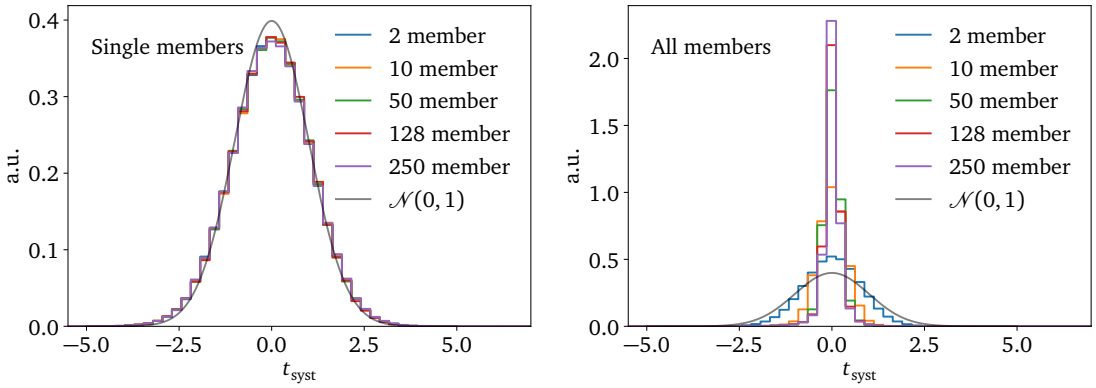

Figure 7.7: Systematic pulls for the repulsive ensemble with a different number of ensemble members. We show the systematic pulls from single ensemble members (left) and the standard averaged prediction of the full ensemble (right).

and after the calibration procedure. The calibration parameter T is estimated using stochastic gradient descent on the full training dataset and evaluated on the test dataset. The loss used for the optimizer is the usual heteroscedastic loss,

$$\mathcal{L}_T(x) = \left\langle \frac{|A_{\text{true}}(x) - \overline{A}(x)|^2}{2\sigma^2(x)T^2} + \log \sigma(x)T \right\rangle_{x \sim D_{\text{train}}} . \tag{7.37}$$

7.4.2 Systematics from network expressivity

Adding artificial noise as the, by definition, dominant uncertainty to our amplitude data immediately bears the question: Where does the systematic uncertainty in the limit of no noise in Fig. 7.6 come from?

In this section, we examine the effect of network size and network expressivity on systematic uncertainty. Fig. 7.9 shows the learned systematics as a function of the noise and for different numbers of hidden layers. Starting with the deterministic network and a heteroscedastic loss, without any additional noise, the median relative systematics

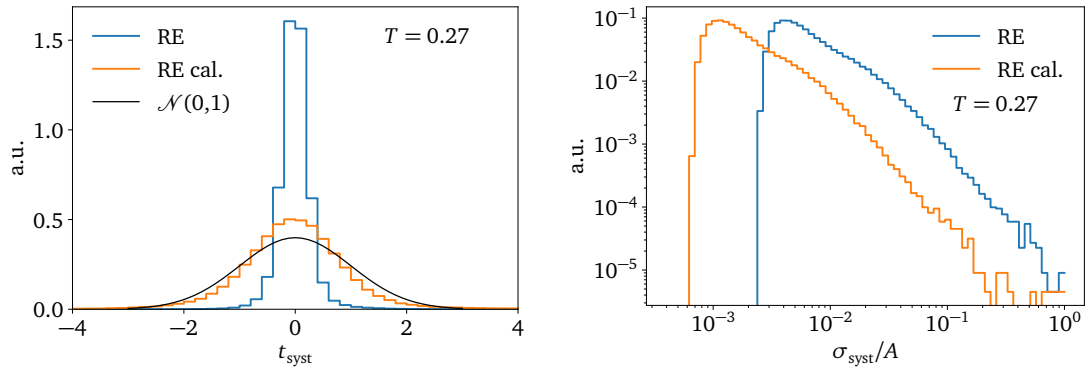

Figure 7.8: Systematic pulls (left) and σ_{syst}/A (right) for the repulsive ensemble before and after the calibration. The fitted value of the temperature parameter is $T = 0.27$.

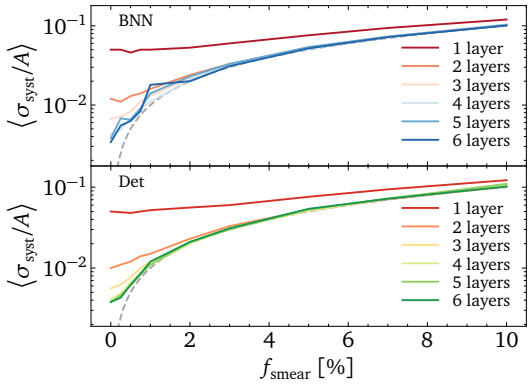

Figure 7.9: Relative uncertainty versus added noise for different numbers of hidden layers. We show results for a deterministic network with heteroscedastic loss and for the BNN. The exact numbers are given in Tab. C.3.

decreases from 5% for one hidden layer to better than 0.5% for five or six hidden layers. On the other hand, we see that training more than three or four hidden layers is starting to be less stable. For just one hidden layer, all noise scenarios are not learned correctly. This indicates that the network is too small to extract the amplitudes and corresponding uncertainty. This improves for two hidden layers, at least down to 2% noise, and for three hidden layers to 0.25% noise. This means that more expressive networks can describe the amplitudes with smaller and smaller noise, finally reaching a systematics plateau around $\langle \sigma_{\text{syst}}/A \rangle = 0.38\%$.

Also, Fig. 7.9 shows the results for repeating the same study for the BNN, which has to separate these joint systematics from the statistical uncertainty, according to Fig. 7.5 and Tab. C.2 at the 0.1% level in the limit of little or no noise. We know that very large BNNs run into stability issues when the Bayesian layers destabilize the training. The reason for this is a too large deterministic network can switch off unused weights by setting them to zero. A BNN can only do this for the mean, while the widths of the network parameter will be driven to the prior hyperparameter. During training, these parameters with zero mean but finite width add unwanted noise. The solution is to only Bayesianize the number of layers needed to express the learned uncertainty. Specifically, we only use a Bayesian layer as the last layer for our networks, which has more than three hidden layers. With this caveat, the BNN results in Fig. 7.9 reproduce the results from the heteroscedastic loss, even with more stable training thanks to the BNN regularization.

Fig. 7.10 displays the systematic pull distributions again, now as a function of the number of hidden layers. All pull distributions are close to the expected unit Gaussian for stochastic sources of the underlying systematics. Given that the systematics we are looking at is the increasing expressive power of the network, this Gaussian distribution is not guaranteed. For one hidden layer, we also show the lowest and highest quantiles in the amplitude size separately. Both of them drive the deviation of the pull distribution from the unit Gaussian, indicating that learning the extreme amplitude values challenges the network expressivity. The situation improves from two hidden layers onward. For large networks, the BNN's pull distributions become slightly too wide, which means the systematic uncertainty is underestimated. This can be explained by the small number of actual Bayesian network weights and the choice of the prior hyperparameter.

In Fig. 7.11, we check the stability of the BNN with four and six hidden layers as a

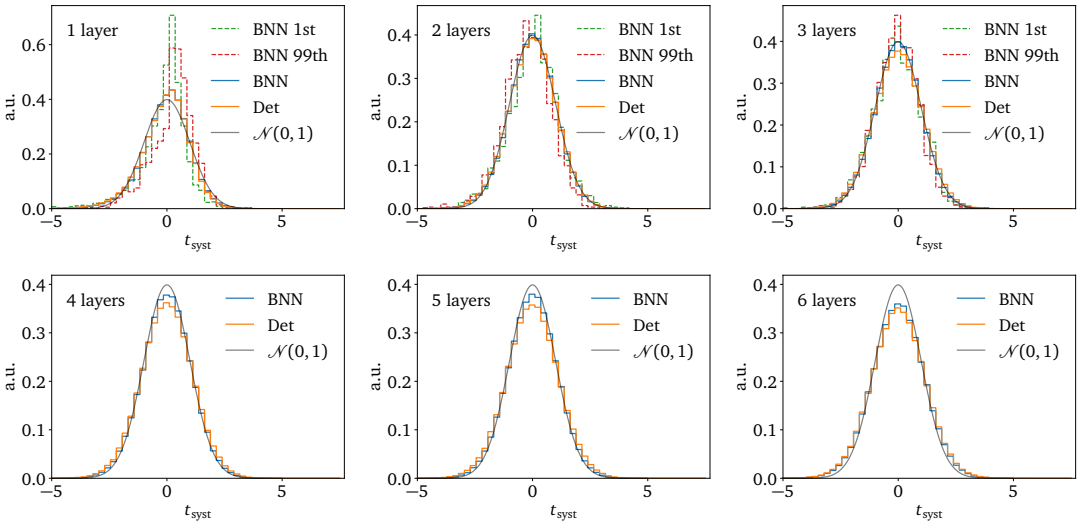

Figure 7.10: Systematic pulls for all test amplitudes without noise, shown for an increasing number of hidden layers. For one and two hidden layers we also show the extreme quantiles of the amplitudes are shown, to identify the failure mode.

function of the prior hyperparameter. While limiting the BNN sampling to the last layer stabilizes the network, the question is if the trained network still samples the entire posterior. To see this, we study the systematic pulls for different values of the prior hyperparameter. We confirm the existence of a broad plateau for this hyperparameter but shift to smaller prior values for more hidden layers. Larger networks with a smaller fraction of Bayesian layers need to be pushed to sample the full statistical uncertainty.

7.4.3 Systematics from symmetries

After identifying added noise and the number of network parameters as the two leading sources of systematic uncertainties, the question is what source of systematics leads to the plateau value for the heteroscedastic network around 0.38% in Fig. 7.9 and Tab. C.3.

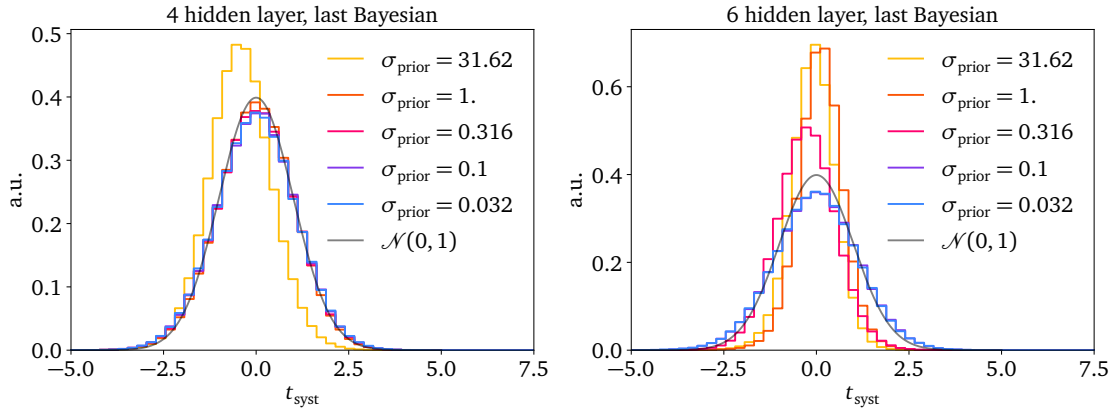

Figure 7.11: Systematic pulls for all test amplitudes without noise, shown for 4 and 6 hidden layers. We use a BNN with different prior widths, and only the last layer is Bayesian.

architecture	# hidden layers					
	1	2	3	4	5	6
$\langle \sigma_{\text{Det}}/A \rangle$ (Tab. C.3)	0.050	0.010	0.0056	0.0041	0.0038	0.0038
$\langle \sigma^{\text{I}}_{\text{Det}}/A \rangle$	0.00380	0.00138	0.00098	0.00086	0.00102	0.00104
$\langle \sigma^{\text{I}}_{\text{Det}}/A \rangle$ float64	—	—	0.00106	—	—	0.00107
$\langle \sigma^{\text{I}}_{\text{Det}}/A \rangle$ float64+leakyReLU	—	—	0.00091	—	—	0.00092
$\langle \sigma^{\text{DS}}_{\text{Det}}/A \rangle$	—	—	0.00019	—	—	—
$\langle \sigma^{\text{DSI}}_{\text{Det}}/A \rangle$	—	—	0.000054	—	—	—
$\langle \sigma^{\text{DSI}}_{\text{Det}}/A \rangle$ with L2-norm	—	—	0.000068	—	—	—
$\langle \sigma^{\text{DSI}}_{\text{Det}}/A \rangle$ 2000 epochs	—	—	0.000039	—	—	—
$\langle \sigma^{\text{DSI}}_{\text{BNN}}/A \rangle$	—	—	0.000070	—	—	—
$\langle \sigma^{\text{DSI}}_{\text{RE}}/A \rangle$	—	—	0.000051	—	—	—

Table 7.1: Comparison of different architectures, starting from a standard heteroscedastic network and adding different features, then turning to deep sets without and with invariants.

To identify this source, we work with more advanced architectures that incorporate the symmetries of our amplitude data. The two key symmetries for LHC amplitudes are Lorentz and permutation invariance [299, 301], where for our simple $(2 \to 3)$ process Lorentz invariance is more relevant. In Tab. 7.1, we document the development when improving the network architecture. We start by considering the Det-I, which includes Mandelstam invariants as additional features, σ^{I}. This leads to a sizeable improvement in the learned uncertainty, which we know tracks the corresponding improvement in accuracy. This improvement exists for all network sizes, with a performance plateau from three to six hidden layers, and for three hidden layers, it reduces the systematics to around 0.1%. Further changes, like higher machine precision or alternative activation functions, do not improve the performance of the standard MLP architecture as deterministic network Det.

In a second step shown in Tab. 7.1, we replace the standard deterministic network with a deep sets architecture designed for amplitude regression, σ^{DS}. We find a significant performance boost to a relative systematic uncertainty around 0.02%. Combining the deep sets architecture with Lorentz invariants defines σ^{DSI}, resulting in another drop in the relative systematic uncertainty to around 0.005%. This level can be stabilized by adding an L2-normalization and training the normalized network for 2000 epochs. For the BNN version of the DSI network, the systematic uncertainty does not quite reach this level but comes extremely close. Notably, when switching to the DSI architecture, we have to train and evaluate the network in double precision to avoid numerical artifacts.

Finally, we look at the systematic pull of the advanced network architectures in Fig. 7.12. The upper panels show the improved accuracy of the networks, now also adding the new Lorentz-equivariant geometric algebra transformer (L-GATr) [299, 301]. This architecture improves the scaling with the number of particles in the final state and is almost on par with the leading DSI, BNN-DSI, and uncalibrated RE-DSI variants. Their accuracy is stable on the 10^{-5} level, with suppressed and symmetric tails. The relative systematic uncertainty is asymmetric, but this is exactly the distribution we expect from a unit-Gaussian pull distribution, reflecting the asymmetric distributions of the training and test amplitudes. All networks, except for the RE variant, have perfectly calibrated systematic uncertainties. This calibration traces the next leading systematics at the 10^{-5} level.

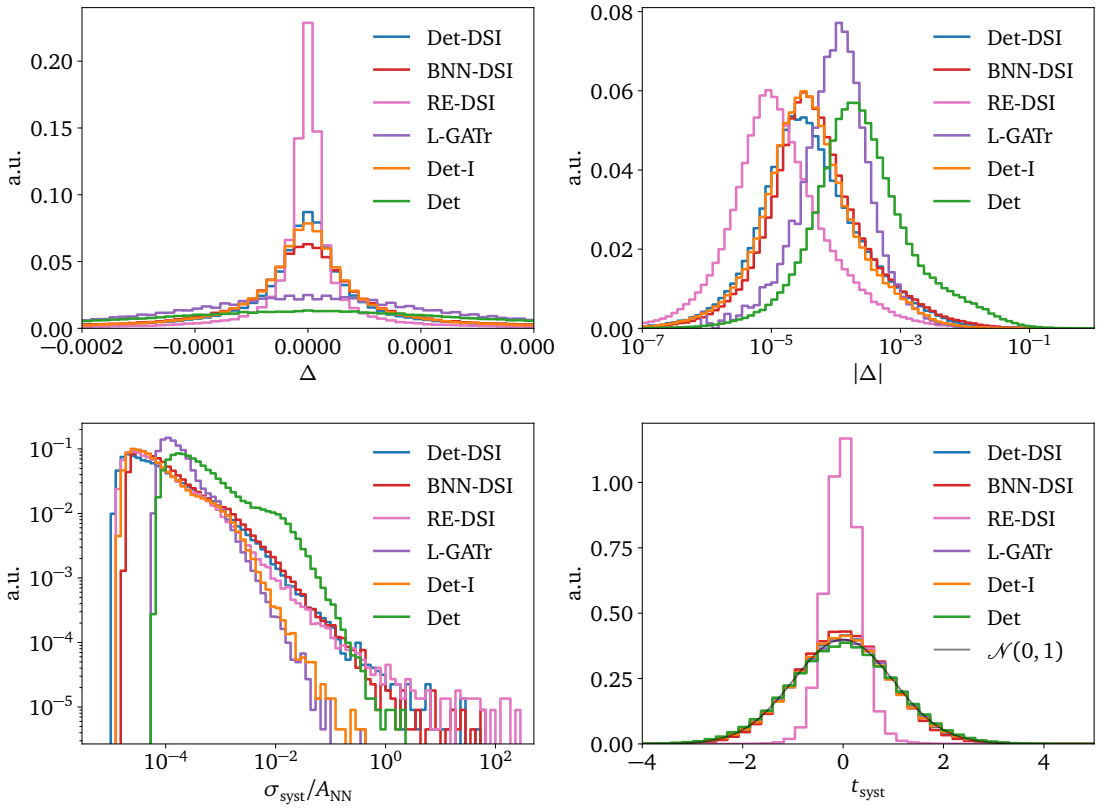

Figure 7.12: Upper: Accuracy on a linear and logarithmic scale for different network architectures using a heteroscedastic loss, as well as the BNN and the RE version of the DSI network. Lower: Relative systematic uncertainties and systematic pulls, indicating the poor calibration of σ_{syst} learned by REs.

7.5 Statistics

The second contribution to the uncertainty arises from limited statistics, and it should vanish in the limit of infinite training data. In this section, we test the statistical pulls defined in Sec. 7.1.4 for both BNN and RE. We start by applying the scaled statistical pull to a large set of amplitudes from the approximate posterior and then turn to the sampled statistical pull. After validating the statistical pull, we confirm that the statistical uncertainties are a sub-leading contribution to the total error for our most precise network.

Scaled statistical uncertainties

We start from a θ-independent pull, as defined in Eq.(7.27), and study its scaling behavior. For these results, we use samples of $N = 512$, which is the largest number of members we can train in parallel for the ensemble. The left panel of Fig. 7.13 shows the pulls for the RE-DSI. The scaling behavior we expect for $t_{\text{stat},M}$ agrees with a standard Gaussian pull for $M \ll N$. As M approaches the number of samples used to estimate σ_{stat}, the correlation between the two variables increases until the scaling of $\sigma_{\text{stat},M}$ does not hold anymore, making the pull narrower. Removing the θ dependence in the A_{true} residuals also does not provide good calibration, as with increasing M, the pulls drift towards

over-confident uncertainties. We observe a similar behavior for the BNN-DSI in the right panel of Fig. 7.13. Thus, we conclude that the sample variance of the mean is not a reliable uncertainty and, for the rest of the section, shifts to the sampled statistical pull with 128 members.

Sampled statistical uncertainties

In Fig. 7.14, we show the sampled statistical pull, defined in Eq.(7.24) for the BNN-DSI and RE-DSI networks from our default training. In the left panel, we use only 10% of the training dataset, and both learned uncertainties are reasonably well calibrated given the limited amount of training data; in the right panel, we use the full training dataset. We see that the statistical uncertainty from the RE-DSI network is calibrated across small and large training datasets. In contrast, the statistical uncertainty of the BNN-DSI network becomes overconfident by roughly a factor of two for the full training dataset.

Finally, we look at the learned uncertainties of the DSI networks as a function of the size of the training dataset. In the left panel of Fig. 7.15, we compare the total relative uncertainties and the accuracies. The asymptotic values for the total uncertainty are similar for all training dataset sizes and coincide with the full dataset. The main difference between the two networks is that the accuracy of the RE-DSI is significantly better because of the assembly. We already know that the main reason for the mismatch between the accuracy and the uncertainty estimate is the poor calibration of the REs for the systematic uncertainties, where the learned uncertainty does not benefit from the ensembling the same way the learned amplitudes do.

The right panel of Fig. 7.15 splits the learned uncertainties into the relative systematic and statistical uncertainties for the BNN and RE. For small training datasets, the uncertainties learned by the two methods behave differently. This is expected for two reasons: First, the two methods approximate the posterior using different approaches with different implicit biases. Second, and more fundamentally, the separation of the total uncertainty into statistical and systematic contributions is not uniquely defined away from the limit of infinite statistics or negligible statistical uncertainties.

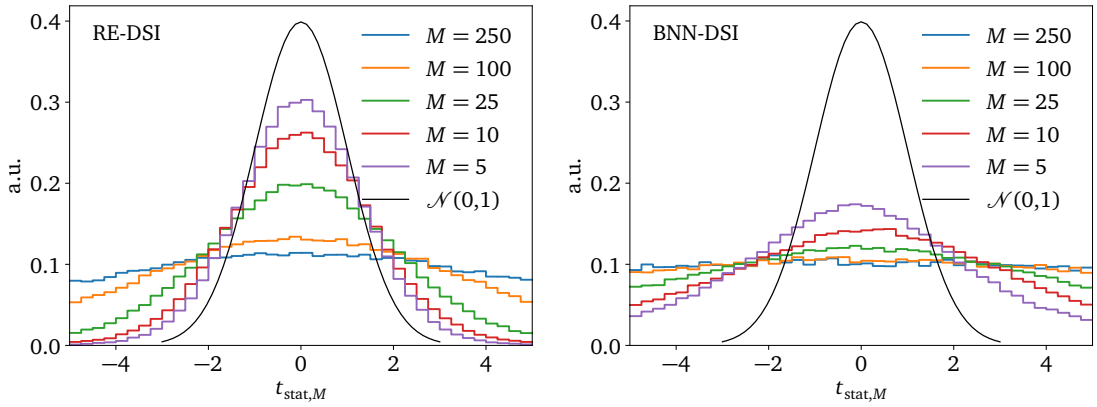

Figure 7.13: Distribution of $t_{\text{stat},M}(x)$, defined in Eq.(7.27), for a test set of amplitudes learned by a RE-DSI (left) and a BNN-DSI (right). The mean $\langle A \rangle$ and the statistical uncertainty σ_{stat} are estimated from 512 amplitudes.

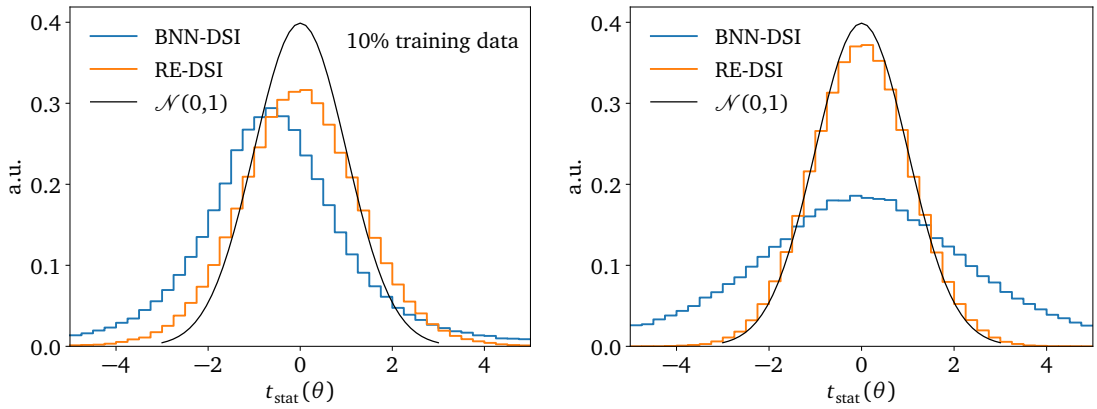

Figure 7.14: Statistical pulls from the exact training data for 10% (left) and all (right) of the training data.

Towards more training data, the systematic uncertainties show a crossing point between BNN and RE, with the ensembles providing smaller uncertainties for large datasets. In this regime, the statistical uncertainties from the BNNs are slightly overconfident, while the statistical RE uncertainties are well calibrated, as we know from Fig. 7.14. On the other hand, the systematic uncertainties from the BNN are perfectly calibrated. This suggests that in splitting the total uncertainty into systematics and statistical parts, the BNN maintains a perfect calibration of the systematics through the heteroscedastic loss at the expense of underestimating the statistical uncertainties by a factor of two.

7.6 Outlook

Fast, accurate, and controlled surrogate amplitudes are a key ingredient to higher-order event generation with future ML event generators. In terms of accuracy, standard MLP architectures have been surpassed, for instance, by a deep-set architecture with Lorentz invariants (DSI) or just by adding an invariant preprocessing; an alternative path might

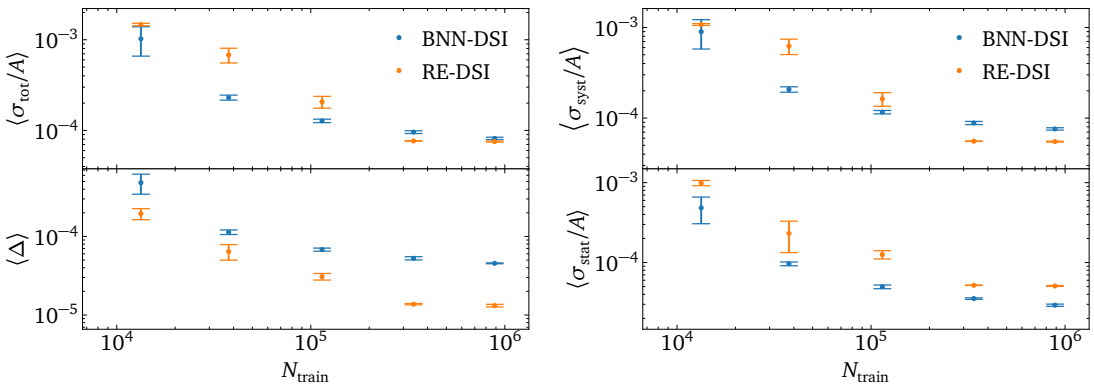

Figure 7.15: Left: Relative total uncertainty for the BNN-DSI and RE-DSI, and relative accuracy. The means and error bars are obtained by averaging over five trainings. Right: Relative uncertainties, split into systematics and statistics, as a function of the size of the training dataset.

be Lorentz-equivariant transformers (L-GATr).

We first studied the impact of activation functions on the accuracy using KANs and learnable activation functions through GroupKANs. While KANs perform worse than a well-chosen fixed activation function, GroupKANs yield comparable performance. This also leads to the conclusion that the chosen activation function has a significant impact on the accuracy of the network predictions, as seen by comparing the ReLU with a GELU activation function for a deterministic network.

For deep networks, appropriate architectures can learn the uncertainties in parallel to the central values for amplitudes over phase space. Heteroscedastic losses in deterministic networks probe systematic uncertainties only, while BNNs and REs, combined with a heteroscedastic loss, track systematic and statistical uncertainties.

For systematic uncertainties, we found that a heteroscedastic loss and the BNN learn well-calibrated uncertainties. We tested this for added noise, network expressivity, and the symmetry implementation in the networks — in decreasing order of the size of the effect on the accuracy and the corresponding uncertainty. REs benefit from their ensemble nature in learning the mean amplitude but not in learning the systematic uncertainty. However, this significant mismatch could be removed through re-calibration. Importantly, we also showed that REs trained without a heteroscedastic loss do not learn any systematic uncertainties.

Statistical uncertainties are currently less relevant for LHC applications because networks are trained on comparably cheap simulations. However, for the DSI architecture, the BNN and the REs indicate that systematic uncertainties are reduced to the current level of statistical uncertainties. Calibrating the statistical uncertainties is conceptually challenging. We find that the BNN estimate is overconfident by roughly a factor of two, while the REs provide a calibrated statistical uncertainty.

Altogether, we have found that for surrogate loop amplitudes, learned uncertainties provide a meaningful way to control the training, identify challenges, and quantify the accuracy of the surrogates. They are key to understanding the improvement in relative accuracy from the percent level for naive networks to the 10^{-5} accuracy level for modern architectures. For the most relevant systematic uncertainties in practice, BNNs are sensitive to a wide range of sources of systematics and provide us with calibrated uncertainty estimates throughout.

Chapter 8

Summary and Outlook

With the beginning of the HL-LHC, the LHC is entering an era of more precise measurements and producing vast amounts of data. To make the most out of this data, we need fast and precise simulations and analysis techniques. Additionally, we have to include a comprehensive treatment of uncertainties. In this thesis, we presented the impact of uncertainties in two global analyses, first using the SMEFT framework and then using a weak-scale EFT approach for EDM measurements. Further, we discussed different ML architectures and their ability to learn uncertainties and provide precise predictions for amplitude surrogates.

In Chapter 4, we performed a SMEFT global analysis including the top, Higgs, and electroweak sectors and studied the implementation of publicly available likelihoods from three ATLAS top-production measurements in SFitter. We benchmarked our results with the provided pull distributions and correlation matrices to justify the assumptions made in SFitter. Additionally, we included a boosted $t\bar{t}$ production measurement in the all-hadronic channel. This boosted measurement was the driving factor for the improved constraints on kinematically enhanced operators. Thus, the distribution was mainly responsible for the improved constraints in the updated top sector analysis compared to the old data set. Including experimental likelihoods enabled us to correlate systematic uncertainties between measurements of the same experiment more precisely. Considering correlations while neglecting theory uncertainties had a negligible impact on the constraints. On the other hand, including theory uncertainties in addition to correlations improved the constraints on the Wilson coefficients. Additionally, comparing the constraints obtained from profiling and marginalization, we observed a slight shift in the central value. This shift occurred because the profiling approach could benefit from the flat distribution of the theory uncertainties by shifting the central value at no cost in the likelihood.

These findings concluded that theory uncertainties currently dominate the constraints on top-sector operators, which is also evident in the global analysis. In contrast to the top sector, the influence of those uncertainties in the Higgs and electroweak sectors is negligible and only visible on operators strongly correlated with the top sector. By improving systematic and statistical uncertainties in future LHC runs, we also have to reduce theory uncertainties to fully benefit from the upcoming data, for example, by improving the event generation chain.

Moreover, we extended the scope of SFitter by including EDM measurements in a low-energy EFT addressing EDMs below the weak scale in Chapter 5. Thus, we chose the hadronic and weak-scale Lagrangians to describe the EDMs at a GeV scale. This choice was motivated by their ability to address the interactions of electrons and nucleons, the

driving forces in EDMs. The hadronic-scale Lagrangian was linked to the weak-scale SMEFT Lagrangian to obtain relations to SMEFT dimension six operators. In total, we included eleven measurements and seven model parameters. By performing this global analysis, we encountered several challenges. First, the experimental results differ in precision, leading to order of magnitude differences between measurements and model parameters. Thus, some highly constraining measurements limited the single-parameter estimations. Because of this, we considered only a subset of parameters, from which we could determine the exact measurements responsible for the constraints. HfF^+ and ThO turned out as leading measurements on the $d_e - C_S^{(0)}$ sub-space, and we could factorize them out since both measurements only constrain d_e and $C_S^{(0)}$. The global analysis found rich correlation patterns in the hadronic sector, including all corresponding parameters. However, when we included the most substantial measurements in the hadronic sector, namely Hg and the neutron measurement, we encountered some narrow correlations. These narrow correlations did not appear in the single-parameter limits, implicating weaker constraints. Next, we included theory uncertainties on the corresponding model parameters. The theory uncertainties had to be combined from various sources using different conventions and calculations to obtain them. These different conventions resulted in highly asymmetrical uncertainties; some included zero as a possible parameter value. By including zero as a possibility, the fit automatically prefers this as the best option, and the corresponding measurement did not contribute to the constraints of the parameter. When including theory uncertainties by a flat distribution, we allowed for a shift in the central value. This behavior was seen in the results of the global analysis by a disagreement in the central values with and without theory uncertainties. Additionally, excluding some measurement parameter pairs changed the impact ordering for some measurements, leading to slightly different correlation patterns. In conclusion, theory uncertainties significantly impact the correlations and constraints obtained from the global analysis using EDMs. Thus, it is necessary to perform such a global analysis to determine all these effects. Furthermore, we have seen that for obtaining more reliable constraints, the precision of the measurements has to be improved, and theory uncertainties must be reduced.

With the demand for more precise measurements and reduced uncertainties, we started investigating the ability of different machine learning architectures to learn uncertainties and determine the precision of their predictions. To fully understand the effect of various networks on the predictions, we started by using KANs to determine the impact of activation functions on the precision of the output. Comparing standard activation functions to different GroupKAN approaches, we found the GELU activation function to be the best compromise between computing time and precision. Next, we introduced a systematic uncertainty to MLPs, BNNs, and REs using a heteroscedastic loss. The BNNs and REs could also provide an estimate of statistical uncertainty. By adding external noise via Gaussian smearing to the amplitudes, we observed that the different architectures could learn external noise as an additional systematic uncertainty. For any external noise exceeding 2% of the amplitude value, noise is the dominant uncertainty source. This behavior was also reflected in the pull distribution of the systematic uncertainty, as all architectures provided perfectly calibrated results for 5% noise. We observed a reasonable estimate and calibration for BNNs for the systematic uncertainties without artificial noise. In contrast to the BNN, the RE benefited from its ensemble nature in precision and in estimating the statistical uncertainties. Additionally, the different architectures obtained uncertainty from the heteroscedastic loss. To better understand this uncertainty from the loss function, we fine-tuned the network expressivity by varying the number of

hidden layers for the MLP and BNN. For the BNN, the prior was varied by three orders of magnitude to ensure stable training and guarantee a prior plateau. Including the kinematic invariants in the input features of the network had the most significant effect in reducing the model uncertainties. This resulted in an overall model uncertainty estimate and precision of the predictions for MLPs, BNNs, and REs of 0.001% or order 10^{-5}. Comparing this number to actual systematics from experiments and theory uncertainties, they are negligibly small.

Overall, we conclude that a reliable and robust uncertainty treatment is key for making the most of future HL-LHC runs and other experimental measurements. As future tasks, amplitude surrogates must be tested on more advanced and complicated processes involving higher-order corrections to ensure precision stability. Further, we want to implement them in the actual event generation chain, for example, as provided by MadGraph5_aMC@NLO [218] or MadNIS [294, 295], to increase the level of precision and speed up compared to the current treatment. The ultimate goal is to perform NLO simulations with the same speed as LO simulations. Also, more precise calculations would benefit more areas in physics, as we have seen in the global analysis of EDM measurements. Again, theory uncertainties are the driving factor of the parameter constraints. They lead to huge differences between the limits obtained from a single parameter estimation and a complete global analysis. In future analyses, correlations can be introduced between measurements and parameters, as in the SMEFT analysis, and their impact can be determined compared to the existing global analysis. Also, different parameterizations can be tested on their correlation patterns and constraints since many different parameterization schemes are available for EDMs. The weak-scale EFT provides a link to SMEFT dimension six operators. Linking both induces the goal of performing a global analysis linking the SMEFT operators used in the top, Higgs, and electroweak sectors to EDM measurements and their parameters.

In conclusion, theory uncertainties are crucial in global analyses and will become even more critical when entering the HL-LHC phase. Thus, we have to develop new frameworks and adapt existing tools with the help of ML architectures. To tackle these challenges, we must provide a fast, precise, flexible, and trustworthy framework for event generation and uncertainty estimation.

Chapter A

Higgs, top, di-boson and electroweak combination

In producing the global analysis of Sec. 4.3.4 and Fig. 4.14, we have combined the top sector from Ref. [1] with the Higgs, di-boson, and electroweak sectors from Ref. [75], taking all data from within these references. The corresponding Wilson coefficients for the Higgs, di-boson, and electroweak sectors in the Warsaw basis are provided in Ch. 2, Eqs. (2.16)-(2.19). The final parameter set thus consists of the 22 operators from the top sector and 20 operators from the Higgs, di-boson, and electroweak sectors and the additional Branching ratio of the Higgs to dark matter agents BR_{inv}.

The numerical values of the boundaries of the 95% CL intervals shown in Fig. 4.14 for all 43 parameters are reported in Table A.1.

Coefficient	Full analysis	Halved theory unc.	Coefficient	Full analysis	Halved theory unc.
$C_{\phi G}$	[-9.25, 6.35]	[-5.56, 5.1]	C_{tG}	[-0.46, 0.35]	[-0.12, 0.19]
$C_{\phi\Box}$	[-1.14, 1.72]	[-1.0, 1.27]	$C_{Qq}^{(18)}$	[-0.29, 0.08]	[-0.21, 0.08]
$C_{u\phi,33}$	[-7.03, 9.11]	[-7.03, 9.11]	$C_{Qq}^{(38)}$	[-0.23, 0.15]	[-0.23, 0.09]
$C_{\phi e} \times 10$	[-2.73, 1.79]	[-2.73, 1.59]	$C_{tq}^{(8)}$	[-0.37, 0.15]	[-0.26, 0.12]
$C_{\phi b} \times 10$	[-11.46, -1.25]	[-10.17, -1.33]	$C_{Qu}^{(8)}$	[-0.49, 0.2]	[-0.43, 0.18]
$C_{\phi D} \times 10$	[-3.48, 5.4]	[-3.33, 5.4]	$C_{Qd}^{(8)}$	[-0.7, 0.4]	[-0.68, 0.33]
$C_{\phi B} \times 10$	[-2.12, 0.63]	[-2.15, 0.63]	$C_{tu}^{(8)}$	[-0.38, 0.1]	[-0.35, 0.1]
$C_{\phi W} \times 10$	[-3.15, 4.29]	[-3.0, 4.28]	$C_{td}^{(8)}$	[-0.42, 0.2]	[-0.4, 0.11]
$C_{\phi WB} \times 10$	[-2.24, 1.64]	[-2.24, 1.26]	$C_{Qq}^{(11)}$	[-0.1, 0.06]	[-0.08, 0.05]
$C_{\phi l}^{(1)} \times 10$	[-1.14, 1.07]	[-1.14, 1.0]	$C_{Qq}^{(31)}$	[-0.08, 0.06]	[-0.06, 0.04]
$C_{\phi l}^{(3)} \times 10$	[-1.38, 0.27]	[-1.36, 0.17]	$C_{tq}^{(1)}$	[-0.08, 0.09]	[-0.07, 0.1]
$C_W \times 20$	[-1.1, 1.2]	[-1.1, 1.12]	$C_{Qu}^{(1)}$	[-0.08, 0.11]	[-0.09, 0.1]
$C_{d\phi,33} \times 20$	[-0.94, 1.51]	[-0.58, 1.36]	$C_{Qd}^{(1)}$	[-0.15, 0.14]	[-0.13, 0.1]
$C_{\phi d} \times 20$	[-2.83, 3.81]	[-2.19, 3.05]	$C_{tu}^{(1)}$	[-0.11, 0.08]	[-0.1, 0.09]
$C_{\phi u} \times 20$	[-1.75, 1.39]	[-1.75, 1.31]	$C_{td}^{(1)}$	[-0.14, 0.12]	[-0.14, 0.11]
$C_{\phi q}^{(3)} \times 20$	[-1.56, 0.84]	[-1.54, 0.8]	$C_{\phi Q}^{(3)}$	[-0.66, 0.32]	[-0.56, 0.23]
$C_{\phi q}^{(1)} \times 20$	[-1.39, 1.08]	[-1.46, 1.08]	C_{tW}	[-0.16, 0.31]	[-0.13, 0.26]
$C_{e\phi,22} \times 100$	[-0.29, 0.58]	[-0.3, 0.58]	$C_{bW}/10$	[-0.17, 0.19]	[-0.12, 0.12]
$C_{e\phi,33} \times 100$	[-1.35, 2.06]	[-1.26, 1.37]	$C_{tZ}/10$	[-0.3, 0.17]	[-0.24, 0.09]
$C_{ll} \times 100$	[-4.61, 0.21]	[-4.51, 0.0]	$C_{\phi Q}^{(1)}/10$	[-0.56, 0.78]	[-0.41, 0.79]
$\mathrm{BR}_{\mathrm{inv}}$	[0, 7.6]	[0, 7.03]	$C_{\phi tb}/10$	[-0.54, 0.48]	[-0.37, 0.36]
			$C_{\phi t}/100$	[-0.2, 0.11]	[-0.17, 0.07]

Table A.1: Numerical values for the 95% CL limits shown in Fig. 4.14. We emphasize that the reduction of the theory uncertainties by a factor two is entirely hypothetical.

Chapter B

Alternative parametrization

In an alternative parametrization, we choose $d^{\rm sr}_{n,p}$ as our model parameters and perform the global SFitter analysis for the set

$$c_j \in \left\{ d_e, C_S^{(0)}, C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr} \right\}, \tag{B.1}$$

rather than Eq.(2.42). To remove the proton EDM parameter we now identify

$$d_p^{\rm sr} \approx -d_n^{\rm sr} \tag{B.2}$$

instead of Eq.(2.41). The linear relations from Eq.(5.13) turn into

$$\begin{aligned} k_{i,S} S_i &= \sum_{c_j \in \{d^{\rm sr}_{n,p}, g_\pi^{(0,1,2)}\}} \alpha_{i,c_j} c_j \\ &\approx k_{i,S} \left[s_{i,n} d_n^{\rm sr} + s_{i,p} d_p^{\rm sr} + \frac{m_N g_A}{F_\pi} \left(a_{i,0} g_\pi^{(0)} + a_{i,1} g_\pi^{(1)} \right) \right] , \end{aligned} \tag{B.3}$$

and the coefficients $a_{i,0}$ and $a_{i,1}$ now include additional terms from the relation between $d_{n,p}$ and $d^{\rm sr}_{n,p}$. In this Appendix we collect versions of all figures from the main body of the paper, derived using the alternative model parameter choice.

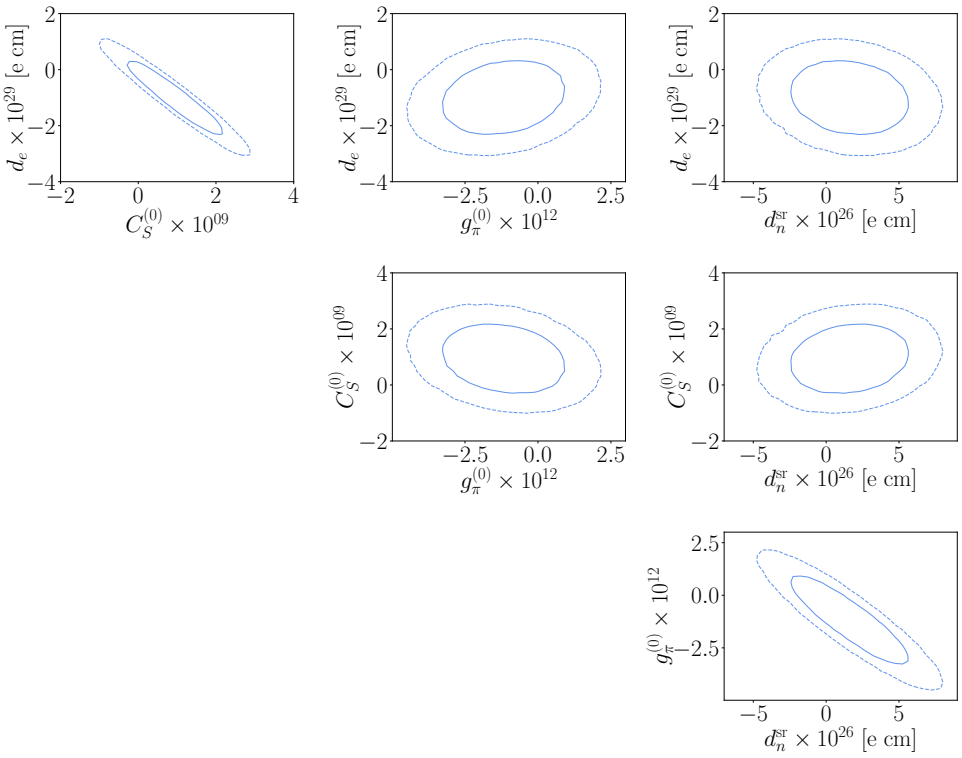

Figure B.1: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n^{\rm sr}\}$, based on all EDM measurements but neglecting theory uncertainties. The ellipses indicate 68% and 95% CL. Figure corresponding to Fig. 5.1 for the $d_{n,p}$ parametrization.

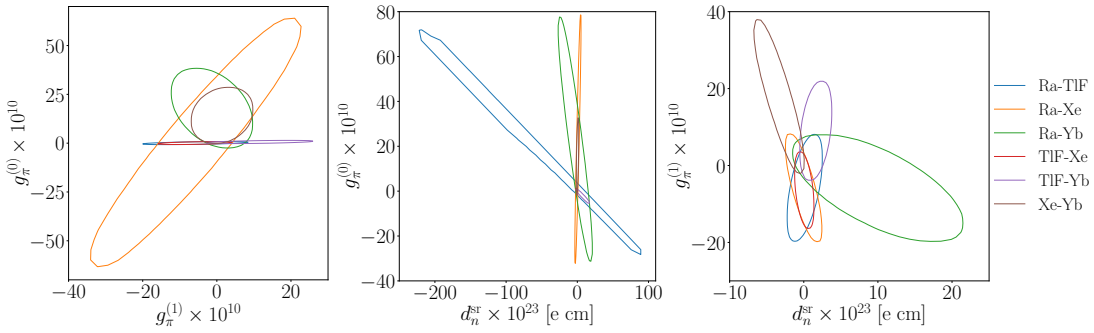

Figure B.2: Correlations from three 2-dimensional analyses in the $\{g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$ parameter space, each based on a different pair of closed-shell EDM measurements, as indicated by the color. The ellipses indicate 68% CL, neglecting theory uncertainties. Figure corresponding to Fig. 5.2 for the $d_{n,p}$ parametrization.

System i	$k_{i,S}$ [cm/fm^3]	$s_{i,n}$ [fm^2]	$s_{i,p}$ [fm^2]
Tl	$-4.2^{+2.1}_{-1.8} \cdot 10^{-18}$ [253]	$0.14^{\pm 0.03}$	$-0.38^{+1.38}_{-0.45}$
Cs	$-9.99^{+2.9}_{-4.1} \cdot 10^{-18}$ [253]		$0.1^{\pm 0.1}$
^{199}Hg	$-2.26^{\pm 0.23} \cdot 10^{-17}$ [345]	$0.6^{+1.33}_{-0.12}$	$0.06^{+0.20}_{-0.01}$
^{129}Xe	$3.62^{\pm 0.25} \cdot 10^{-18}$ [345]	$0.63^{+0.16}_{-0.12}$	$0.14^{\pm 0.03}$
^{171}Yb	$-2.10^{+0.22}_{-0.0} \cdot 10^{-17}$ [252, 346]	$0.54^{+0.13}_{-0.11}$	$0.054^{+0.016}_{-0.014}$
^{225}Ra	$-8.5^{+0.25}_{-0.3} \cdot 10^{-17}$ [252, 346, 347]	$0.63^{+0.16}_{-0.12}$	$0.14^{+0.04}_{-0.03}$
TlF	$-4.59^{\pm 0.41} \cdot 10^{-13}$ [345]	$0.14^{\pm 0.03}$	$-0.38^{+1.38}_{-0.45}$
	$a_{i,0}$ [e fm^3]	$a_{i,1}$ [e fm^3]	$a_{i,2}$ [e fm^3]
Tl	$0.113^{+0.017}_{-0.008}$	$-0.004^{+0.0}_{-0.006}$	$-0.226^{+0.044}_{-0.03}$ [348]
Cs	$-0.006^{+0.0}_{-0.074}$	$-0.02^{\pm 0.01}$	$-0.04^{+0.0}_{-0.017}$ [349]
^{199}Hg	$0.01^{+0.4}_{-0.005}$	$0.02^{+0.07}_{-0.05}$	$0.02^{+0.04}_{-0.01}$ [24]
^{129}Xe	$-0.008^{+0.003}_{-0.042}$	$0.006^{+0.044}_{-0.003}$	$-0.009^{+0.004}_{-0.091}$ [349]
^{171}Yb	$0.01^{+0.02}_{-0.0}$	$0.02^{+0.034}_{-0.027}$	$0.02^{+0.04}_{-0.01}$ [350]
^{225}Ra	$-1.5^{+0.5}_{-4.5}$	6^{+18}_{-2}	-4^{+3}_{-11} [351]
TlF	$0.113^{+0.017}_{-0.008}$	$-0.004^{+0.0}_{-0.006}$	$-0.226^{+0.044}_{-0.03}$ [348]

Table B.1: Quantities used as inputs for computing the hadronic coefficients in Tab.s 5.3 and B.2. Ranges are inferred from the distribution of literature values, and should be taken as indicative; signs are adapted as necessary to match our conventions. In some cases where the valence nucleon is n (respectively p) literature values are not available for the $s_{i,p}$ (respectively, $s_{i,n}$). For these cases we estimate based on the spin fractions of Tab. 5.2, as done previously for near-spherical nuclei within the shell model [352]. For some discussion of these estimations, and in particular difficulties associated with $s_{\mathrm{Tl},p}$ and $s_{\mathrm{Hg},n}$ see Ref. [346] and references therein. We do not report the coefficient $s_{\mathrm{Cs},p}$ separately from $\alpha_{\mathrm{Cs},p}$, since the latter includes also contributions from a nuclear magnetic quadrupole moment [353, 354]. This should be borne in mind if using Eq. 5.13 for ^{133}Cs or nuclei with spin $I > 1/2$; see also [250, 355] for relating a magnetic quadrupole moment to our coefficients $a_{i,j}$. References given for $a_{i,2}$ indicate the sources for central values for all three coefficients $a_{i,j}$; the ranges are inferred from the broader distribution of literature values, see e.g. [24] for a related discussion.

System i	α_{i,d_e}	$\alpha_{i,C_S^{(0)}}$ [e cm]	$\alpha_{i,C_P^{(0)}}$ [e cm]	$\alpha_{i,C_T^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(1)}}$ [e cm]	$\alpha_{i,d_n^{\rm sr}}$	$\alpha_{i,d_p^{\rm sr}}$
n	–	–	–	–	$1.38^{\pm 0.02} \cdot 10^{-14}$	$2.73^{\pm 0.02} \cdot 10^{-16}$	1	(-1)
^{205}Tl	$-558^{\pm 28}$ [249]	$-6.77^{\pm 0.34} \cdot 10^{-18}$	$1.4^{+2.5}_{-0.8} \cdot 10^{-19}$	$8.8^{+4.0}_{-1.2} \cdot 10^{-21}$	$-6.74^{+4.85}_{-5.12} \cdot 10^{-18}$	$2.20^{+6.53}_{-1.44} \cdot 10^{-19}$	$-5.75^{+3.59}_{-3.77} \cdot 10^{-6}$	$1.61^{+3.35}_{-7.56} \cdot 10^{-5}$
^{133}Cs	$123^{\pm 4}$	$7.80^{+0.2}_{-0.8} \cdot 10^{-19}$	$-1.4^{+0.8}_{-2.2} \cdot 10^{-20}$	$1.7^{\pm 0.4} \cdot 10^{-20}$	–	–	–	–
^{199}Hg	$11.6^{+10}_{-18} \cdot 10^{-3}$ [253, 257]	$-1.26^{+0.7}_{-1.6} \cdot 10^{-21}$	$6.6^{+2.4}_{-1.6} \cdot 10^{-23}$	$-6.4^{+3}_{-4} \cdot 10^{-21}$	$-4.70^{+3.13}_{-19.4} \cdot 10^{-18}$	$-6.13^{+16.7}_{-25.7} \cdot 10^{-18}$	$-1.36^{+0.39}_{-3.45} \cdot 10^{-4}$	$-1.36^{+0.39}_{-5.11} \cdot 10^{-5}$
^{129}Xe	$-8^{+0}_{-8} \cdot 10^{-4}$ [253, 257]	$-2.1^{+1.2}_{-2.5} \cdot 10^{-22}$	$1.7^{+0.5}_{-0.4} \cdot 10^{-23}$	$1.24^{+0.78}_{-0.61} \cdot 10^{-21}$	$-1.53^{+3.08}_{-24.5} \cdot 10^{-19}$	$2.98^{+24.5}_{-1.71} \cdot 10^{-19}$	$2.29^{+0.77}_{-0.58} \cdot 10^{-5}$	$4.89^{+1.64}_{-1.25} \cdot 10^{-6}$
^{171}Yb	$1.44^{+165}_{-4.5} \cdot 10^{-3}$ [253]	$-1.31^{+0.55}_{-1.06} \cdot 10^{-21}$	$4.93^{+3.54}_{-1.55} \cdot 10^{-23}$	$-3.68^{+1.86}_{-2.43} \cdot 10^{-21}$	$-4.21^{+0.94}_{-6.52} \cdot 10^{-18}$	$-5.70^{+7.77}_{-10.4} \cdot 10^{-18}$	$-1.13^{+0.32}_{-0.28} \cdot 10^{-4}$	$-1.13^{+0.32}_{-0.28} \cdot 10^{-5}$
^{225}Ra	$-5.4^{+134}_{-2.0} \cdot 10^{-2}$ [253]	$1.13^{+2.94}_{-0.51} \cdot 10^{-20}$	$-7.63^{+2.05}_{-3.88} \cdot 10^{-22}$	$-4.5^{+2.0}_{-2.5} \cdot 10^{-20}$	$1.72^{+5.73}_{-0.67} \cdot 10^{-15}$	$-6.89^{+2.66}_{-22.9} \cdot 10^{-15}$	$-5.38^{+1.20}_{-1.58} \cdot 10^{-4}$	$-1.19^{+0.27}_{-0.35} \cdot 10^{-4}$
TlF	$1.36^{+0.36}_{-0.32} \cdot 10^{3}$ [253]	$1.44^{+0.8}_{-0.5} \cdot 10^{-16}$	$1.51^{+2.2}_{-5.6} \cdot 10^{-18}$	$1.87^{+0.19}_{-0.17} \cdot 10^{-15}$	$-7.36^{+2.51}_{-2.60} \cdot 10^{-13}$	$2.40^{+4.93}_{-0.53} \cdot 10^{-14}$	$-6.28^{+2.01}_{-1.72} \cdot 10^{-1}$	$1.76^{+2.41}_{-6.76}$
HfF^+	1	$9.17^{\pm 0.06} \cdot 10^{-21}$	–	–	–	–	–	–
ThO	1	$1.51^{+0}_{-0.2} \cdot 10^{-20}$	–	–	–	–	–	–
YbF	1	$8.99^{\pm 0.70} \cdot 10^{-21}$	–	–	–	–	–	–
	$\eta_{i,d_e}^{(m)} \left[\frac{\text{mrad}}{\text{s } e \text{ cm}}\right]$	$k_{i,C_S}^{(m)} \left[\frac{\text{mrad}}{\text{s}}\right]$	α_{i,C_P}	α_{i,C_T}	$\alpha_{i,g_\pi^{(0)}}$	$\alpha_{i,g_\pi^{(1)}}$	$\alpha_{i,d_n^{\rm sr}}$	$\alpha_{i,d_p^{\rm sr}}$
HfF^+	$3.49^{\pm 0.14} \cdot 10^{28}$ [251, 258–261]	$3.2^{+0.1}_{-0.2} \cdot 10^{8}$ [251, 258, 259]	–	–	–	–	–	–
ThO	$-1.21^{+0.05}_{-0.39} \cdot 10^{29}$ [251, 262–264]†	$-1.82^{+0.42}_{-0.27} \cdot 10^{9}$ [251, 262, 264–266]†	–	–	–	–	–	–
YbF	$-1.96^{\pm 0.15} \cdot 10^{28}$ [251, 265–268]	$-1.76^{\pm 0.2} \cdot 10^{8}$ [251, 265–267]	–	–	–	–	–	–

Table B.2: Central values and theory uncertainties for the α-parameters defined in Eq.(5.1), now using Eq.(5.5) to treat $d_{n,p}^{\rm sr}$ as model parameters and assuming $d_n^{\rm sr} = -d_p^{\rm sr}$ in lieu of Eq.(2.41). The coefficients $\alpha_{i,g_\pi^{(0,1)}}$ are accordingly modified via Eq.(5.13), and differ from those in Tab.5.3. Here a "$-$" means that we neglect the dependence in our global analysis. †There appears to be an overall sign issue in the coefficients reported for ThO in Table 4 of Ref. [251]. The values for Tl and Cs of $\alpha_{i,C_T^{(0)}}$ are estimated by simple analytical calculations [250], and the uncertainties quoted here are estimated as approximately twice those arising from the relevant hadronic matrix elements.

System i	d_e [e cm]	$C_S^{(0)}$	$C_P^{(0)}$	$C_T^{(0)}$
Tl	$(7.2 \pm 7.7) \cdot 10^{-28}$	$(5.9 \pm 6.4) \cdot 10^{-8}$	$(-2.9 \pm 3.1) \cdot 10^{-6}$	$(-4.5 \pm 4.9) \cdot 10^{-5}$
Cs	$(-1.5 \pm 5.6) \cdot 10^{-26}$	$(-2.3 \pm 8.9) \cdot 10^{-6}$	$(1.3 \pm 5.0) \cdot 10^{-4}$	$(-1.1 \pm 4.1) \cdot 10^{-4}$
^{199}Hg	$(1.9 \pm 2.7) \cdot 10^{-28}$	$(-1.7 \pm 2.5) \cdot 10^{-9}$	$(3.3 \pm 4.7) \cdot 10^{-8}$	$(-3.4 \pm 4.9) \cdot 10^{-10}$
^{129}Xe	$(2.2 \pm 2.3) \cdot 10^{-25}$	$(8.4 \pm 8.7) \cdot 10^{-7}$	$(-1.0 \pm 1.1) \cdot 10^{-5}$	$(-1.4 \pm 1.5) \cdot 10^{-7}$
^{171}Yb	$(-4.7 \pm 3.6) \cdot 10^{-24}$	$(5.2 \pm 4.0) \cdot 10^{-6}$	$(-1.4 \pm 1.1) \cdot 10^{-4}$	$(1.8 \pm 1.4) \cdot 10^{-6}$
^{225}Ra	$(-0.7 \pm 1.1) \cdot 10^{-22}$	$(3.5 \pm 5.3) \cdot 10^{-4}$	$(-5.2 \pm 7.9) \cdot 10^{-3}$	$(-0.9 \pm 1.3) \cdot 10^{-4}$
TlF	$(-1.3 \pm 2.1) \cdot 10^{-26}$	$(-1.2 \pm 2.0) \cdot 10^{-7}$	$(-1.1 \pm 1.9) \cdot 10^{-5}$	$(-0.9 \pm 1.6) \cdot 10^{-8}$
HfF^+	$(-1.3 \pm 2.1) \cdot 10^{-30}$	$(-1.4 \pm 2.3) \cdot 10^{-10}$		
ThO	$(4.3 \pm 4.0) \cdot 10^{-30}$	$(2.8 \pm 2.7) \cdot 10^{-10}$		
YbF	$(-2.4 \pm 5.9) \cdot 10^{-28}$	$(-2.7 \pm 6.6) \cdot 10^{-8}$		

	$g_\pi^{(0)}$	$g_\pi^{(1)}$	d_n^{sr}	d_p^{sr}
n	$(0 \pm 8.1) \cdot 10^{-13}$	$(0 \pm 4.1) \cdot 10^{-11}$	$(0 \pm 1.1) \cdot 10^{-26}$	$(0 \pm 1.1) \cdot 10^{-26}$
Tl	$(5.9 \pm 6.4) \cdot 10^{-8}$	$(-1.8 \pm 2.0) \cdot 10^{-6}$	$(7.0 \pm 7.5) \cdot 10^{-20}$	$(-2.5 \pm 2.7) \cdot 10^{-20}$
^{199}Hg	$(-4.7 \pm 6.6) \cdot 10^{-13}$	$(-3.6 \pm 5.1) \cdot 10^{-13}$	$(-1.6 \pm 2.3) \cdot 10^{-26}$	$(-1.6 \pm 2.3) \cdot 10^{-25}$
^{129}Xe	$(1.2 \pm 1.2) \cdot 10^{-9}$	$(5.9 \pm 6.1) \cdot 10^{-10}$	$(-7.7 \pm 7.9) \cdot 10^{-24}$	$(-3.6 \pm 3.7) \cdot 10^{-23}$
^{171}Yb	$(1.6 \pm 1.2) \cdot 10^{-9}$	$(1.2 \pm 0.9) \cdot 10^{-9}$	$(6.0 \pm 4.6) \cdot 10^{-23}$	$(6.0 \pm 4.6) \cdot 10^{-22}$
^{225}Ra	$(2.3 \pm 3.5) \cdot 10^{-9}$	$(-5.8 \pm 8.7) \cdot 10^{-10}$	$(-0.7 \pm 1.1) \cdot 10^{-20}$	$(-3.4 \pm 5.0) \cdot 10^{-20}$
TlF	$(2.3 \pm 3.9) \cdot 10^{-11}$	$(-0.7 \pm 1.2) \cdot 10^{-9}$	$(2.7 \pm 4.6) \cdot 10^{-23}$	$(-1.0 \pm 1.6) \cdot 10^{-23}$

Table B.3: Single-parameter ranges allowed by each of the EDM measurements given in Tab. 5.1, using the coefficients from Tab. B.2.

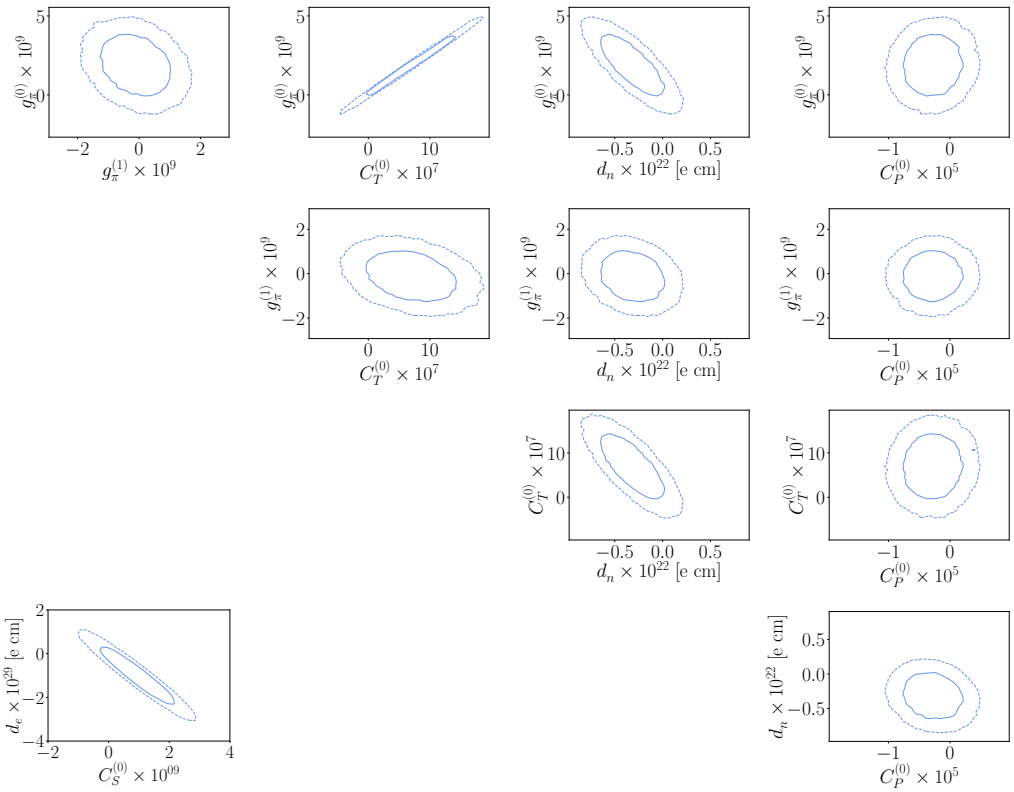

Figure B.3: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$ and the factorized $d_e - C_S^{(0)}$ plane from Fig. 5.1. We ignore the neutron and Hg measurements, which induce narrow correlation patterns in the 5-dimensional parameters space and do not affect the profiled 2-dimensional correlations. The ellipses indicate 68% and 95% CL, neglecting theory uncertainties. Figure corresponding to Fig. 5.3 for the $d_{n,p}$ parametrization.

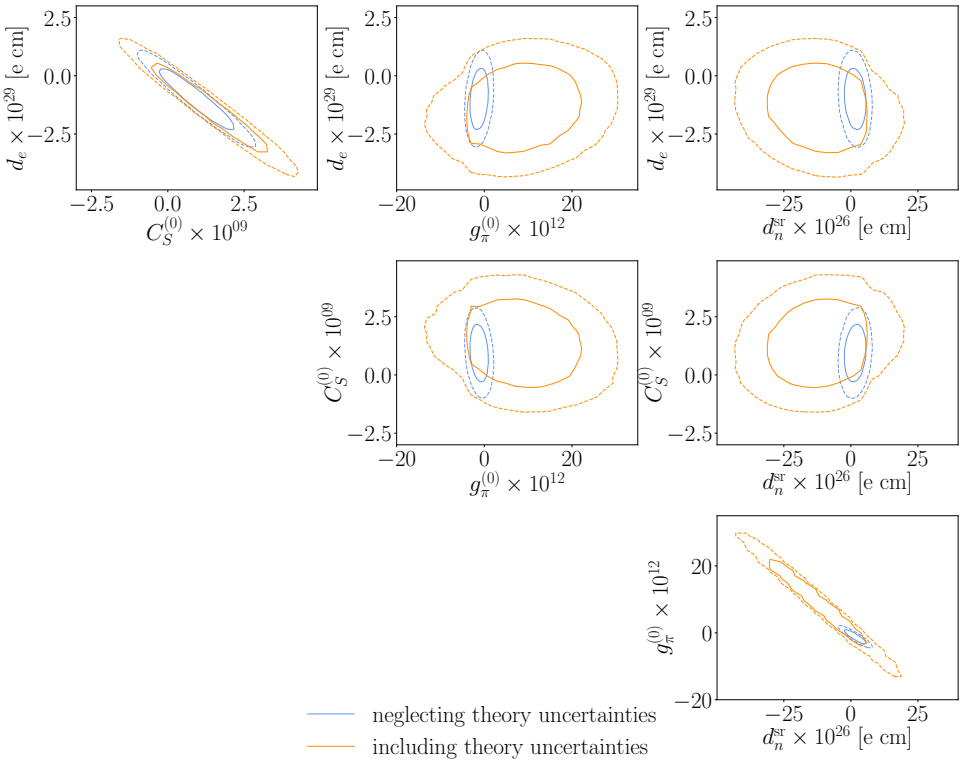

Figure B.4: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n^{\rm sr}\}$. The orange curves show the effect of theory uncertainties on the results of Fig. 5.1. The ellipses indicate 68% and 95% CL. Figure corresponding to Fig. 5.4 for the $d_{n,p}$ parametrization.

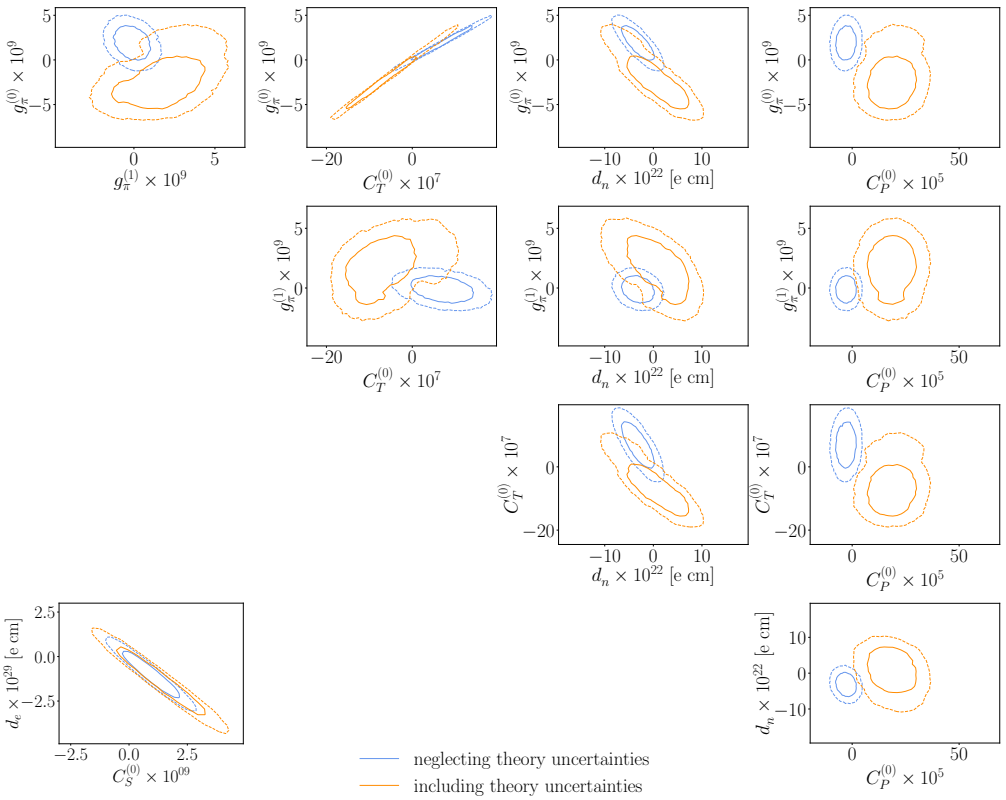

Figure B.5: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$, and the factorized $d_e - C_S^{(0)}$ plane from Fig. 5.4. The orange curves show the effect of theory uncertainties on the results of Fig 5.3. The ellipses indicate 68% and 95% CL. Figure corresponding to Fig. 5.5 for the $d_{n,p}$ parametrization.

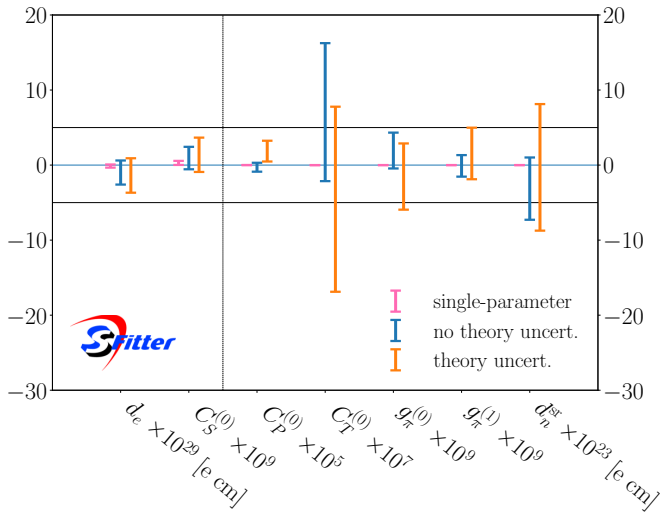

Figure B.6: 68% CL constraints from the global EDM analysis on the parameters of the hadronic-scale Lagrangian. We show (i) hugely over-constrained single-parameter ranges allowed by the best available measurement; (ii) over-optimistic allowed ranges for profiled single parameters, ignoring theory uncertainties; (iii) allowed ranges for profiled single parameters including experimental and theory uncertainties. Figure corresponding to Fig. 5.6 for the $d_{n,p}$ parametrization.

Chapter C

Learning Uncertainties: Additional material and hyperparameters

Hyperparameters

This section provides the hyperparameters for the training of an MLP and a DS(I) combination.

Parameter	MLP	DS(I)
Size of latent rep.	-	64
Activation function	ReLU	GELU
Number of layers	3	3
Hidden nodes	128	128
Batch size	1024	1024
Scheduler	One cycle	Cosine
Max learning rate	10^{-3}	10^{-3}
Number of epochs	1000	1000

Table C.1: Network and training parameters of the MLP/DSI.

Summary tables

A detailed table for all networks versus noise split into different uncertainty contributions: In the last rows of Tab. C.2, we show the total uncertainties learned by the two networks, by defining the square sum of the statistic and systematics. The BNN and RE results are similar. Just out of interest, we also show the uncertainty from the RE in a setup where we do not add the additional heteroscedastic loss. We see that it overestimates the total uncertainty without added noise and does not track the systematics from added noise at all. This indicates that the REs without heteroscedastic uncertainty are not well-suited to extract a systematic uncertainty like added noise.

		0%	0.25%	0.5%	0.75%	1%	2%	3%	5%	7%	10%
$\langle\sigma_{\text{het}}/A\rangle$	25	0.0054	0.0061	0.0073	0.0094	0.012	0.021	0.030	0.051	0.072	0.104
	50	0.0047	0.0053	0.0068	0.0090	0.011	0.021	0.030	0.051	0.072	0.103
	75	0.0047	0.0053	0.0067	0.0089	0.011	0.021	0.030	0.051	0.073	0.104
$\langle\sigma_{\text{syst, BNN}}/A\rangle$	25	0.0067	0.0069	0.0082	0.010	0.013	0.022	0.032	0.051	0.072	0.103
	50	0.0053	0.0060	0.0076	0.010	0.012	0.021	0.031	0.052	0.072	0.103
	75	0.0054	0.0059	0.0076	0.010	0.012	0.022	0.032	0.052	0.073	0.104
$\langle\sigma_{\text{syst, RE}}/A\rangle$	25	0.0054	0.0061	0.0076	0.0095	0.012	0.021	0.031	0.050	0.070	0.101
	50	0.0045	0.0054	0.0070	0.0090	0.011	0.021	0.030	0.050	0.071	0.101
	75	0.0045	0.0052	0.0069	0.0090	0.011	0.021	0.030	0.050	0.070	0.101
$\langle\sigma_{\text{stat, BNN}}/A\rangle$	25	0.0012	0.0013	0.0015	0.0018	0.0020	0.0028	0.0039	0.0061	0.0071	0.0084
	50	0.0012	0.0012	0.0014	0.0018	0.0019	0.0027	0.0037	0.0059	0.0072	0.0085
	75	0.0012	0.0013	0.0016	0.0019	0.0022	0.0031	0.0041	0.0066	0.0081	0.010
$\langle\sigma_{\text{stat, RE}}/A\rangle$	25	0.0057	0.0059	0.0061	0.0065	0.0066	0.0076	0.0088	0.0106	0.012	0.015
	50	0.0046	0.0050	0.0053	0.0056	0.0057	0.0068	0.0078	0.0096	0.011	0.015
	75	0.0045	0.0048	0.0052	0.0055	0.0055	0.0066	0.0077	0.0094	0.011	0.013
$\langle\sigma_{\text{tot, BNN}}/A\rangle$	25	0.0068	0.0071	0.0083	0.011	0.013	0.022	0.032	0.052	0.072	0.104
	50	0.0055	0.0061	0.0077	0.010	0.012	0.021	0.032	0.052	0.072	0.103
	75	0.0055	0.0061	0.0078	0.010	0.012	0.022	0.032	0.053	0.073	0.105
$\langle\sigma_{\text{tot, RE}}/A\rangle$	25	0.0078	0.0085	0.0098	0.012	0.013	0.022	0.032	0.051	0.072	0.104
	50	0.0065	0.0073	0.0088	0.011	0.013	0.022	0.031	0.051	0.072	0.102
	75	0.0064	0.0071	0.0087	0.011	0.013	0.022	0.031	0.051	0.071	0.102
$\langle\sigma_{\text{MSE, RE}}/A\rangle$	25	0.0081	0.0080	0.0082	0.0082	0.0084	0.0089	0.0096	0.0111	0.013	0.015
	50	0.0074	0.0073	0.0074	0.0075	0.0077	0.0081	0.0087	0.0101	0.011	0.014
	75	0.0073	0.0073	0.0073	0.0074	0.0075	0.0079	0.0085	0.0098	0.011	0.014

Table C.2: Learned uncertainties as a function of added noise, in terms of the median relative uncertainty for three amplitude quantiles. We use three hidden layers and show the learned systematic, statistical, and total uncertainties. For the latter, we also show the RE trained only with an MSE loss instead of a full heteroscedastic loss.

	# layers	0%	0.25%	0.5%	0.75%	1%	2%	3%	5%	7%	10%
$\langle\sigma_{\text{Det}}/A\rangle$	1	0.050	0.049	0.048	0.050	0.052	0.056	0.060	0.076	0.094	0.122
	2	0.010	0.011	0.012	0.014	0.015	0.023	0.033	0.052	0.071	0.102
	3	0.0056	0.0062	0.0077	0.010	0.012	0.021	0.031	0.051	0.072	0.104
	4	0.0041	0.0048	0.0066	0.0086	0.011	0.020	0.031	0.052	0.073	0.105
	5	0.0038	0.0046	0.0061	0.0083	0.011	0.021	0.030	0.053	0.072	0.110
	6	0.0038	0.0043	0.0062	0.0085	0.012	0.021	0.031	0.054	0.072	0.102
$\langle\sigma_{\text{syst, BNN}}/A\rangle$	1	0.050	0.050	0.046	0.050	0.050	0.053	0.060	0.076	0.094	0.120
	2	0.012	0.011	0.013	0.014	0.016	0.024	0.033	0.053	0.073	0.103
	3	0.0067	0.0071	0.0083	0.011	0.013	0.022	0.032	0.052	0.073	0.104
$\langle\sigma_{\text{syst, BNN}}/A\rangle$	4	0.0043	0.0049	0.0068	0.0090	0.011	0.021	0.033	0.051	0.073	0.103
	5	0.0038	0.0068	0.0065	0.0091	0.014	0.023	0.033	0.054	0.073	0.103
	6	0.0034	0.0055	0.0063	0.0084	0.018	0.020	0.031	0.052	0.072	0.101

Table C.3: Learned uncertainties as a function of added noise and the number of hidden layers, each with 128 dimensions. For the BNN with 4 or more hidden layers only the last layer is Bayesian, and the prior hyperparameter is $\sigma_{\text{prior}} = 0.316$ for 4 and 5 hidden layers and $\sigma_{\text{prior}} = 0.1$ for 6 hidden layers.

	0%	0.25%	0.5%	0.75%	1%	2%
$\langle\sigma_{\text{syst, DSI RE}}/A\rangle$	$5.1{\cdot}10^{-5}$	0.00249	0.00498	0.00753	0.0100	0.0205
$\langle\sigma_{\text{syst, DSI BNN}}/A\rangle$	$7.0{\cdot}10^{-5}$	0.00251	0.00499	0.00754	0.0100	0.0201
$\langle\sigma_{\text{stat, DSI RE}}/A\rangle$	$4.8{\cdot}10^{-5}$	0.00014	0.00025	0.00042	0.00068	0.00136
$\langle\sigma_{\text{stat, DSI BNN}}/A\rangle$	$2.3{\cdot}10^{-5}$	0.00016	0.00026	0.00070	0.00083	0.0014
$\langle\sigma_{\text{tot, DSI RE}}/A\rangle$	$7.0{\cdot}10^{-5}$	0.00250	0.00500	0.00756	0.0100	0.0205
$\langle\sigma_{\text{tot, DSI BNN}}/A\rangle$	$7.4{\cdot}10^{-5}$	0.00451	0.00506	0.00757	0.0101	0.0201

Table C.4: Learned uncertainties as a function of added noise, in terms of the median relative uncertainty. We use the DSI network.

Acknowledgments

First of all, I am grateful to my advisor, Tilman Plehn, for his support over the last four years and for not only accepting me as a master- but also for keeping me as a PhD student. I would like to thank him for introducing me to machine learning, giving me the opportunity to work on many different and interesting projects, and sharing some life lessons and wisdom with us – some were more helpful than others.

Furthermore, I would like to thank Monica Dunford for refereeing this thesis and Skyler Degenkolb and Björn Malte Schäfer for completing my examination committee.

I gratefully acknowledge support from the International Max Planck Research School "Precision Tests of Fundamental Symmetries".

Moreover, I would like to thank everyone I have worked and collaborated with over the past three years. Special thanks to Maeve Madigan for all the discussions about SMEFT and for pointing out the flaws in SFitter, to Skyler Degenkolb for introducing me to EDMs and the three kinds of fun, to Ramon Winterhalder for always taking time to answer my questions, discussing the results and coming up with new ideas, for the fun nights in the London Pubs and morning run at CERN and to Henning Bahl for the great collaboration in the Learned Uncertainties project and the fruitful discussions about machine learning. Furthermore, I am grateful for all the other collaborators I haven't mentioned yet, Nikita Schmal, Tanmoy Modak, Margarete Mühlleitner, and Luigi Favaro.

Furthermore, I would also like to thank Ramon Winterhalder, Theo Heimel, Henning Bahl, Lennart Röver, and Lorenz Vogel for proofreading this thesis and making it readable and nice-looking.

Most importantly, I would like to thank all the current and former group members and people of the Philosophenweg for making the last three years so enjoyable. In no particular order, I would like to thank Luigi Favaro for inviting Theo and me to visit him and his family during our summer vacation and for the fun ride from Pittsburgh to New York with lots of AJR songs; to Theo Heimel for all the trips and adventures we experienced together, supporting and always being there for me; to Nikita Schmal for being part of the SFitter crew and introducing happiness via happy fitter; to Maeve Madigan for our trip to Brookhaven and your fascination for planes; to Henning Bahl for all the discussions about physics and machine learning; to Benedikt Schosser for reminding me of the importance of a good work-life balance; to Rebecca Kuntz for all the coffee breaks; to Lennart Röver for initiating spike ball matches; to Lorenz Vogel for discussions about politics, meetings at the Marstall and paying for other peoples drinks too often; to Jonas Spinner for the visits of Handball matches and the early morning chats in London due to false fire alarms; as well as Victor Breso, Anja Butter, Giovanni De Crescenzo, Florian Ernst, Nathan Hütsch, Claudius Krause, Ayodele Ore, Sofia Palacios

Schweitzer, Antoine Petitjean, Sebastian Pitz, Daniel Schiller, Paula Schuchard, Sophia Vent, Javier Villadamigo, Heinrich von Campe, and everyone else I forgot to mention.

I am also grateful for the people supporting me outside the institute. Special thanks to Charlotte Nägle for always being there for me despite your busy life, to Judith Bünte and Ann-Kristin Lüvelsmeyer for our regular Monday evening calls, to Isabel Onkelbach for introducing me to road cycling and all the game nights we had, and to Holger Hartwig and Ingo Schmidt for the advice of "Go with the flow".

Finally, I would like to thank my family, especially my parents, for their support throughout the years. It wouldn't be possible without you!

Bibliography

[1] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, *Staying on Top of SMEFT-Likelihood Analyses*, SciPost Phys. **18** (2025) 108, arXiv:2312.12502 [hep-ph].

[2] S. Degenkolb, N. Elmer, T. Modak, M. Mühlleitner, and T. Plehn, *A Global View of the EDM Landscape*, (3, 2024) , arXiv:2403.02052 [hep-ph].

[3] H. Bahl, N. Elmer, L. Favaro, M. Haußmann, T. Plehn, and R. Winterhalder, *Accurate Surrogate Amplitudes with Calibrated Uncertainties*, (12, 2024) , arXiv:2412.12069 [hep-ph].

[4] ATLAS, 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, arXiv:1207.7214 [hep-ex].

[5] CMS, 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, arXiv:1207.7235 [hep-ex].

[6] ATLAS, G. Aad *et al.*, *Standard Model Summary Plots February 2022*, (2022) .

[7] ATLAS, G. Aad *et al.*, *SUSY March 2023 Summary Plot Update*, (2023) .

[8] ATLAS, G. Aad *et al.*, *Summary Plots for Heavy Particle Searches and Long-lived Particle Searches - March 2023*, (2023) .

[9] ATLAS, G. Aad *et al.*, *Dark matter summary plots for s-channel, 2HDM+a and Dark Higgs models*, (2022) .

[10] ATLAS, G. Aad *et al.*, *Standard Model Summary Plots June 2024*, (2024) .

[11] ATLAS, G. Aad *et al.*, *Summary plot for the interpretation of heavy Higgs boson searches in the Georgi-Machacek model*, tech. rep., CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-008.

[12] ATLAS, G. Aad *et al.*, *Summary of Diboson Resonance Searches at the ATLAS experiment using full run-2 data*, (2023) .

[13] ATLAS, G. Aad *et al.*, *Summary of Exotic Higgs Boson Decays from the ATLAS Experiment*, (2021) .

[14] CMS, A. Hayrapetyan *et al.*, *Search for light long-lived particles decaying to displaced jets in proton–proton collisions at* $\sqrt{s} = 13.6$ *TeV*, Rept. Prog. Phys. **88** (2025) 3, 037801, arXiv:2409.10806 [hep-ex].

[15] CMS, A. Hayrapetyan *et al.*, *Search for a heavy resonance decaying into a Z and a Higgs boson in events with an energetic jet and two electrons, two muons, or missing transverse momentum in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, JHEP **02** (2025) 089, arXiv:2411.00202 [hep-ex].

[16] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32.

[17] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Progress in electroweak baryogenesis*, Ann. Rev. Nucl. Part. Sci. **43** (1993) 27, arXiv:hep-ph/9302210.

[18] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry*, Mod. Phys. Lett. A **9** (1994) 795, arXiv:hep-ph/9312215.

[19] P. Huet and E. Sather, *Electroweak baryogenesis and standard model CP violation*, Phys. Rev. D **51** (1995) 379, arXiv:hep-ph/9404302.

[20] M. B. Gavela, M. Lozano, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry. Part 1: Zero temperature*, Nucl. Phys. B **430** (1994) 345, arXiv:hep-ph/9406288.

[21] M. B. Gavela, P. Hernandez, J. Orloff, O. Pene, and C. Quimbay, *Standard model CP violation and baryon asymmetry. Part 2: Finite temperature*, Nucl. Phys. B **430** (1994) 382, arXiv:hep-ph/9406289.

[22] A. Riotto and M. Trodden, *Recent progress in baryogenesis*, Ann. Rev. Nucl. Part. Sci. **49** (1999) 35, arXiv:hep-ph/9901362.

[23] M. Pospelov and A. Ritz, *Electric dipole moments as probes of new physics*, Annals Phys. **318** (2005) 119, arXiv:hep-ph/0504231.

[24] J. Engel, M. J. Ramsey-Musolf, and U. van Kolck, *Electric Dipole Moments of Nucleons, Nuclei, and Atoms: The Standard Model and Beyond*, Prog. Part. Nucl. Phys. **71** (2013) 21, arXiv:1303.2371 [nucl-th].

[25] N. Yamanaka, B. K. Sahoo, N. Yoshinaga, T. Sato, K. Asahi, and B. P. Das, *Probing exotic phenomena at the interface of nuclear and particle physics with the electric dipole moments of diamagnetic atoms: A unique window to hadronic and semi-leptonic CP violation*, Eur. Phys. J. A **53** (2017) 3, 54, arXiv:1703.01570 [hep-ph].

[26] T. Chupp, P. Fierlinger, M. Ramsey-Musolf, and J. Singh, *Electric dipole moments of atoms, molecules, nuclei, and particles*, Rev. Mod. Phys. **91** (2019) 1, 015001, arXiv:1710.02504 [physics.atom-ph].

[27] W. Dekens, J. de Vries, M. Jung, and K. K. Vos, *The phenomenology of electric dipole moments in models of scalar leptoquarks*, JHEP **01** (2019) 069, arXiv:1809.09114 [hep-ph].

[28] D. McKeen, M. Pospelov, and A. Ritz, *Modified Higgs branching ratios versus CP and lepton flavor violation*, Phys. Rev. D **86** (2012) 113004, arXiv:1208.4597 [hep-ph].

[29] S. Dwivedi, D. K. Ghosh, B. Mukhopadhyaya, and A. Shivaji, *Constraints on CP-violating gauge-Higgs operators*, Phys. Rev. D **92** (2015) 9, 095015, arXiv:1505.05844 [hep-ph].

[30] Y. T. Chien, V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Direct and indirect constraints on CP-violating Higgs-quark and Higgs-gluon interactions*, JHEP **02** (2016) 011, arXiv:1510.00725 [hep-ph].

[31] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Is there room for CP violation in the top-Higgs sector?*, Phys. Rev. D **94** (2016) 1, 016002, arXiv:1603.03049 [hep-ph].

[32] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Constraining the top-Higgs sector of the Standard Model Effective Field Theory*, Phys. Rev. D **94** (2016) 3, 034031, arXiv:1605.04311 [hep-ph].

[33] E. Mereghetti, J. de Vries, W. H. Hockings, C. M. Maekawa, and U. van Kolck, *The Electric Dipole Form Factor of the Nucleon in Chiral Perturbation Theory to Sub-leading Order*, Phys. Lett. B **696** (2011) 97, arXiv:1010.4078 [hep-ph].

[34] G. Apollinari, O. Brüning, T. Nakamoto, and L. Rossi, *High Luminosity Large Hadron Collider HL-LHC*, CERN Yellow Rep. (2015) 5, 1, arXiv:1705.08830 [physics.acc-ph].

[35] J. M. Campbell *et al.*, *Event generators for high-energy physics experiments*, SciPost Phys. **16** (2024) 5, 130, arXiv:2203.11110 [hep-ph].

[36] J. Bendavid, *Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks*, (6, 2017) , arXiv:1707.00028 [hep-ph].

[37] M. D. Klimek and M. Perelstein, *Neural Network-Based Approach to Phase Space Integration*, SciPost Phys. **9** (2020) 053, arXiv:1810.11509 [hep-ph].

[38] I.-K. Chen, M. D. Klimek, and M. Perelstein, *Improved neural network Monte Carlo simulation*, SciPost Phys. **10** (2021) 1, 023, arXiv:2009.07819 [hep-ph].

[39] C. Gao, J. Isaacson, and C. Krause, *i-flow: High-dimensional Integration and Sampling with Normalizing Flows*, Mach. Learn. Sci. Tech. **1** (1, 2020) 045023, arXiv:2001.05486 [physics.comp-ph].

[40] E. Bothmann, T. Janßen, M. Knobbe, T. Schmale, and S. Schumann, *Exploring phase space with Neural Importance Sampling*, SciPost Phys. **8** (1, 2020) 069, arXiv:2001.05478 [hep-ph].

[41] C. Gao, S. Höche, J. Isaacson, C. Krause, and H. Schulz, *Event Generation with Normalizing Flows*, Phys. Rev. D **101** (2020) 7, 076002, arXiv:2001.10028 [hep-ph].

[42] K. Danziger, T. Janßen, S. Schumann, and F. Siegert, *Accelerating Monte Carlo event generation – rejection sampling using neural network event-weight estimates*, SciPost Phys. **12** (9, 2022) 164, arXiv:2109.11964 [hep-ph].

[43] T. Janßen, D. Maître, S. Schumann, F. Siegert, and H. Truong, *Unweighting multijet event generation using factorisation-aware neural networks*, SciPost Phys. **15** (2023) 3, 107, arXiv:2301.13562 [hep-ph].

[44] E. Bothmann, T. Childers, W. Giele, F. Herren, S. Hoeche, J. Isaacsson, M. Knobbe, and R. Wang, *Efficient phase-space generation for hadron collider event simulation*, SciPost Phys. **15** (2023) 169, arXiv:2302.10449 [hep-ph].

[45] F. Bishara and M. Montull, *Machine learning amplitudes for faster event generation*, Phys. Rev. D **107** (2023) 7, L071901, arXiv:1912.11055 [hep-ph].

[46] S. Badger and J. Bullock, *Using neural networks for efficient evaluation of high multiplicity scattering amplitudes*, JHEP **06** (2020) 114, arXiv:2002.07516 [hep-ph].

[47] J. Aylett-Bullock, S. Badger, and R. Moodie, *Optimising simulations for diphoton production at hadron colliders using amplitude neural networks*, JHEP **08** (2021) 066, arXiv:2106.09474 [hep-ph].

[48] D. Maître and H. Truong, *A factorisation-aware Matrix element emulator*, JHEP **11** (2021) 066, arXiv:2107.06625 [hep-ph].

[49] S. Badger, A. Butter, M. Luchmann, S. Pitz, and T. Plehn, *Loop amplitudes from precision networks*, SciPost Phys. Core **6** (2023) 034, arXiv:2206.14831 [hep-ph].

[50] A. Dersy, M. D. Schwartz, and X. Zhang, *Simplifying Polylogarithms with Machine Learning*, Int. J. Data Sci. Math. Sci. **1** (2024) 2, 135, arXiv:2206.04115 [cs.LG].

[51] V. Bresó, G. Heinrich, V. Magerya, and A. Olsson, *Interpolating amplitudes*, arXiv:2412.09534 [hep-ph].

[52] D. J. C. Mackay, *Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks*, Netw. Comput. Neural Syst. **6** (1995) 3, 469.

[53] R. M. Neal, *Bayesian learning for neural networks*. PhD thesis, Toronto, 1995.

[54] Y. Gal, *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge, 2016.

[55] A. Kendall and Y. Gal, *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?*, Proc. NIPS (2017) , arXiv:1703.04977 [cs.CV].

[56] F. D'Angelo and V. Fortuin, *Repulsive deep ensembles are bayesian*, CoRR (2021) , arXiv:2106.11642.

[57] L. Röver, B. M. Schäfer, and T. Plehn, *PINNferring the Hubble Function with Uncertainties*, arXiv:2403.13899 [astro-ph.CO].

[58] H. Bahl, V. Bresó, G. De Crescenzo, and T. Plehn, *Advancing Tools for Simulation-Based Inference*, arXiv:2410.07315 [hep-ph].

[59] A. V. Manohar, *Introduction to Effective Field Theories*, (4, 2018) , arXiv:1804.05863 [hep-ph].

[60] A. A. Petrov and A. E. Blechman, *Effective Field Theories*. WSP, 2016.

[61] C. P. Burgess, *Introduction to Effective Field Theory*, Ann. Rev. Nucl. Part. Sci. **57** (2007) 329, arXiv:hep-th/0701053.

[62] B. Gripaios, *Lectures on Effective Field Theory*, (6, 2015) , arXiv:1506.05039 [hep-ph].

[63] D. B. Kaplan, *Five lectures on effective field theory*, arXiv preprint nucl-th/0510023 (2005) .

[64] R. Penco, *An Introduction to Effective Field Theories*, (6, 2020) , arXiv:2006.16285 [hep-th].

[65] S. Willenbrock and C. Zhang, *Effective Field Theory Beyond the Standard Model*, Ann. Rev. Nucl. Part. Sci. **64** (2014) 83, arXiv:1401.0470 [hep-ph].

[66] A. Falkowski, *Lectures on SMEFT*, Eur. Phys. J. C **83** (2023) 7, 656.

[67] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, Phys. Rept. **793** (2019) 1, arXiv:1706.08945 [hep-ph].

[68] B. Henning, X. Lu, and H. Murayama, *How to use the Standard Model effective field theory*, JHEP **01** (2016) 023, arXiv:1412.1837 [hep-ph].

[69] A. Falkowski, *Effective field theory approach to LHC Higgs data*, Pramana **87** (2016) 3, 39, arXiv:1505.00046 [hep-ph].

[70] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10** (2010) 085, arXiv:1008.4884 [hep-ph].

[71] S. Weinberg, *Baryon and Lepton Nonconserving Processes*, Phys. Rev. Lett. **43** (1979) 1566.

[72] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003.

[73] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology*, JHEP **04** (2014) 159, arXiv:1312.2014 [hep-ph].

[74] K. Hagiwara, S. Ishihara, R. Szalapski, and D. Zeppenfeld, *Low-energy effects of new interactions in the electroweak boson sector*, Phys. Rev. D **48** (1993) 2182.

[75] I. Brivio, S. Bruggisser, N. Elmer, E. Geoffray, M. Luchmann, and T. Plehn, *To profile or to marginalize - A SMEFT case study*, SciPost Phys. **16** (2024) 1, 035, arXiv:2208.08454 [hep-ph].

[76] G. Durieux, F. Maltoni, and C. Zhang, *Global approach to top-quark flavor-changing interactions*, Phys. Rev. D **91** (2015) 7, 074017, arXiv:1412.7166 [hep-ph].

[77] C. Degrande, F. Maltoni, J. Wang, and C. Zhang, *Automatic computations at next-to-leading order in QCD for top-quark flavor-changing neutral processes*, Phys. Rev. D **91** (2015) 034024, arXiv:1412.5594 [hep-ph].

[78] D. Buarque Franzosi and C. Zhang, *Probing the top-quark chromomagnetic dipole moment at next-to-leading order in QCD*, Phys. Rev. D **91** (2015) 11, 114010, arXiv:1503.08841 [hep-ph].

[79] I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou, S. Westhoff, and C. Zhang, *O new physics, where art thou? A global search in the top sector*, JHEP **02** (2020) 131, arXiv:1910.03606 [hep-ph].

[80] C. W. Murphy, *Dimension-8 operators in the Standard Model Effective Field Theory*, JHEP **10** (2020) 174, arXiv:2005.00059 [hep-ph].

[81] K. K. Vos, H. W. Wilschut, and R. G. E. Timmermans, *Symmetry violations in nuclear and neutron β decay*, Rev. Mod. Phys. **87** (2015) 1483, arXiv:1509.04007 [hep-ph].

[82] P. D. Group, *CP Violation in K0 Decays*, Prog. Theor. Exp. Phys. **2022** (2022) 083C01.

[83] P. D. Group, *CKM Quark-Mixing Matrix*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[84] J. R. Ellis, J. S. Lee, and A. Pilaftsis, *Electric Dipole Moments in the MSSM Reloaded*, JHEP **10** (2008) 049, arXiv:0808.1819 [hep-ph].

[85] Y. Li, S. Profumo, and M. Ramsey-Musolf, *A Comprehensive Analysis of Electric Dipole Moment Constraints on CP-violating Phases in the MSSM*, JHEP **08** (2010) 062, arXiv:1006.1440 [hep-ph].

[86] K. Cheung, T.-J. Hou, J. S. Lee, and E. Senaha, *Higgs Mediated EDMs in the Next-to-MSSM: An Application to Electroweak Baryogenesis*, Phys. Rev. D **84** (2011) 015002, arXiv:1102.5679 [hep-ph].

[87] S. Inoue, M. J. Ramsey-Musolf, and Y. Zhang, *CP-violating phenomenology of flavor conserving two Higgs doublet models*, Phys. Rev. D **89** (2014) 11, 115023, arXiv:1403.4257 [hep-ph].

[88] J. de Vries, E. Mereghetti, R. G. E. Timmermans, and U. van Kolck, *The Effective Chiral Lagrangian From Dimension-Six Parity and Time-Reversal Violation*, Annals Phys. **338** (2013) 50, arXiv:1212.0990 [hep-ph].

[89] C. M. Maekawa, E. Mereghetti, J. de Vries, and U. van Kolck, *The Time-Reversal- and Parity-Violating Nuclear Potential in Chiral Effective Theory*, Nucl. Phys. A **872** (2011) 117, arXiv:1106.6119 [nucl-th].

[90] Y. V. Stadnik and V. V. Flambaum, *Nuclear spin-dependent interactions: Searches for WIMP, Axion and Topological Defect Dark Matter, and Tests of Fundamental Symmetries*, Eur. Phys. J. C **75** (2015) 3, 110, arXiv:1408.2184 [hep-ph].

[91] J. C. Berengut, V. V. Flambaum, and E. M. Kava, *Search for variation of fundamental constants and violations of fundamental symmetries using isotope comparisons*, Phys. Rev. A **84** (2011) 042510, arXiv:1109.1893 [physics.atom-ph].

[92] Flavour Lattice Averaging Group (FLAG), Y. Aoki *et al.*, *FLAG Review 2021*, Eur. Phys. J. C **82** (2022) 10, 869, arXiv:2111.09849 [hep-lat].

[93] HERMES, A. Airapetian *et al.*, *Precise determination of the spin structure function g(1) of the proton, deuteron and neutron*, Phys. Rev. D **75** (2007) 012007, arXiv:hep-ex/0609039.

[94] A. Anselm, V. Bunakov, V. Gudkov, and N. Uraltsev, *On the neutron electric dipole moment in the weinberg cp-violation model*, Physics Letters B **152** (1985) 1, 116.

[95] M. Shifman, A. Vainshtein, and V. Zakharov, *Remarks on higgs-boson interactions with nucleons*, Physics Letters B **78** (1978) 4, 443.

[96] S. M. Barr, *T and P odd electron - nucleon interactions and the electric dipole moments of large atoms*, Phys. Rev. D **45** (1992) 4148.

[97] M. Bauer and T. Plehn, *Yet Another Introduction to Dark Matter: The Particle Physics Approach*, vol. 959 of *Lecture Notes in Physics*. Springer, 2019. arXiv:1705.01987 [hep-ph].

[98] R. Lafaye, T. Plehn, and D. Zerwas, *SFITTER: SUSY parameter analysis at LHC and LC*, arXiv:hep-ph/0404282.

[99] R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Supersymmetry*, Eur. Phys. J. C **54** (2008) 617, arXiv:0709.3985 [hep-ph].

[100] C. Adam, J.-L. Kneur, R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Unification*, Eur. Phys. J. C **71** (2011) 1520, arXiv:1007.2190 [hep-ph].

[101] E. Turlay, R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Supersymmetry with Heavy Scalars*, J. Phys. G **38** (2011) 035003, arXiv:1011.0759 [hep-ph].

[102] D. López-Val, T. Plehn, and M. Rauch, *Measuring extended Higgs sectors as a consistent free couplings model*, JHEP **10** (2013) 134, arXiv:1308.1979 [hep-ph].

[103] A. Butter, T. Plehn, M. Rauch, D. Zerwas, S. Henrot-Versillé, and R. Lafaye, *Invisible Higgs Decays to Hooperons in the NMSSM*, Phys. Rev. D **93** (2016) 015011, arXiv:1507.02288 [hep-ph].

[104] S. Henrot-Versillé, R. Lafaye, T. Plehn, M. Rauch, D. Zerwas, S. Plaszczynski, B. Rouillé d'Orfeuil, and M. Spinelli, *Constraining Supersymmetry using the relic density and the Higgs boson*, Phys. Rev. D **89** (2014) 5, 055017, arXiv:1309.6958 [hep-ph].

[105] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas, and M. Duhrssen, *Measuring the Higgs Sector*, JHEP **08** (2009) 009, arXiv:0904.3866 [hep-ph].

[106] M. Klute, R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Higgs Couplings from LHC Data*, Phys. Rev. Lett. **109** (2012) 101801, arXiv:1205.2699 [hep-ph].

[107] M. Bauer, A. Butter, J. Gonzalez-Fraile, T. Plehn, and M. Rauch, *Learning from a Higgs-like scalar resonance*, Phys. Rev. D **95** (2017) 5, 055011, arXiv:1607.04562 [hep-ph].

[108] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, SciPost Phys. **6** (2019) 6, 064, arXiv:1812.07587 [hep-ph].

[109] A. Butter, O. J. P. Éboli, J. Gonzalez-Fraile, M. C. Gonzalez-Garcia, T. Plehn, and M. Rauch, *The Gauge-Higgs Legacy of the LHC Run I*, JHEP **07** (2016) 152, arXiv:1604.03105 [hep-ph].

[110] T. Corbett, O. J. P. Eboli, D. Goncalves, J. Gonzalez-Fraile, T. Plehn, and M. Rauch, *The Higgs Legacy of the LHC Run I*, JHEP **08** (2015) 156, arXiv:1505.05516 [hep-ph].

[111] A. Biekötter, D. Gonçalves, T. Plehn, M. Takeuchi, and D. Zerwas, *The global Higgs picture at 27 TeV*, SciPost Phys. **6** (2019) 2, 024, arXiv:1811.08401 [hep-ph].

[112] A. Biekötter, *Know its limits: A global view on Higgs couplings at the LHC*. PhD thesis, Heidelberg U., 2019.

[113] A. Butter, *Global Fits for New Physics at the LHC and Beyond*. PhD thesis, U. Heidelberg (main), 2017.

[114] 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) 694-706, 289.

[115] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 1, 60.

[116] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943) 3, 426.

[117] A. A. Markov, *Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga*, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya **15** (1906) 135.

[118] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, The Journal of Chemical Physics **21** (06, 1953) 1087.

[119] W. K. Hastings, *Monte carlo sampling methods using markov chains and their applications*, Biometrika **57** (1970) 1, 97.

[120] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd ed., 2013.

[121] Z. Hioki and K. Ohkuma, *Latest constraint on nonstandard top-gluon couplings at hadron colliders and its future prospect*, Phys. Rev. D **88** (2013) 017503, arXiv:1306.5387 [hep-ph].

[122] A. Buckley, C. Englert, J. Ferrando, D. J. Miller, L. Moore, M. Russell, and C. D. White, *Global fit of top quark effective theory to data*, Phys. Rev. D **92** (2015) 9, 091501, arXiv:1506.08845 [hep-ph].

[123] A. Buckley, C. Englert, J. Ferrando, D. J. Miller, L. Moore, M. Russell, and C. D. White, *Constraining top quark effective theory in the LHC Run II era*, JHEP **04** (2016) 015, arXiv:1512.03360 [hep-ph].

[124] N. P. Hartland, F. Maltoni, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector*, JHEP **04** (2019) 100, arXiv:1901.05965 [hep-ph].

[125] R. Aoude, E. Madge, F. Maltoni, and L. Mantani, *Quantum SMEFT tomography: Top quark pair production at the LHC*, Phys. Rev. D **106** (2022) 5, 055007, arXiv:2203.05619 [hep-ph].

[126] D. Barducci *et al.*, *Interpreting top-quark LHC measurements in the standard-model effective field theory*, arXiv:1802.07237 [hep-ph].

[127] F. Maltoni, L. Mantani, and K. Mimasu, *Top-quark electroweak interactions at high energy*, JHEP **10** (2019) 004, arXiv:1904.05637 [hep-ph].

[128] R. Aoude, H. El Faham, F. Maltoni, and E. Vryonidou, *Complete SMEFT predictions for four top quark production at hadron colliders*, JHEP **10** (2022) 163, arXiv:2208.04962 [hep-ph].

[129] C. Degrande, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation*, JHEP **10** (2018) 005, arXiv:1804.07773 [hep-ph].

[130] H. E. Faham, F. Maltoni, K. Mimasu, and M. Zaro, *Single top production in association with a WZ pair at the LHC in the SMEFT*, JHEP **01** (2022) 100, arXiv:2111.03080 [hep-ph].

[131] S. Bißmann, J. Erdmann, C. Grunwald, G. Hiller, and K. Kröninger, *Constraining top-quark couplings combining top-quark and $\boldsymbol{B}$ decay observables*, Eur. Phys. J. C **80** (2020) 2, 136, arXiv:1909.13632 [hep-ph].

[132] R. Aoude, T. Hurth, S. Renner, and W. Shepherd, *The impact of flavour data on global fits of the MFV SMEFT*, JHEP **12** (2020) 113, arXiv:2003.05432 [hep-ph].

[133] S. Bruggisser, R. Schäfer, D. van Dyk, and S. Westhoff, *The Flavor of UV Physics*, JHEP **05** (2021) 257, arXiv:2101.07273 [hep-ph].

[134] S. Bruggisser, D. van Dyk, and S. Westhoff, *Resolving the flavor structure in the MFV-SMEFT*, JHEP **02** (2023) 225, arXiv:2212.02532 [hep-ph].

[135] R. Bartocci, A. Biekötter, and T. Hurth, *A global analysis of the SMEFT under the minimal MFV assumption*, JHEP **05** (2024) 074, arXiv:2311.04963 [hep-ph].

[136] G. Durieux, A. Irles, V. Miralles, A. Peñuelas, R. Pöschl, M. Perelló, and M. Vos, *The electro-weak couplings of the top and bottom quarks — Global fit and future prospects*, JHEP **12** (2019) 98, arXiv:1907.10619 [hep-ph]. [Erratum: JHEP 01, 195 (2021)].

[137] F. Garosi, D. Marzocca, A. R. Sánchez, and A. Stanzione, *Indirect constraints on top quark operators from a global SMEFT analysis*, JHEP **12** (2023) 129, arXiv:2310.00047 [hep-ph].

[138] 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 [hep-ph].

[139] SMEFiT, 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 [hep-ph].

[140] J. Gao, M. Gao, T. J. Hobbs, D. Liu, and X. Shen, *Simultaneous CTEQ-TEA extraction of PDFs and SMEFT parameters from jet and $t\bar{t}$ data*, JHEP **05** (2023) 003, arXiv:2211.01094 [hep-ph].

[141] Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. Morales Alvarado, J. Rojo, and M. Ubiali, *The top quark legacy of the LHC Run II for PDF and SMEFT analyses*, JHEP **05** (2023) 205, arXiv:2303.06159 [hep-ph].

[142] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Can New Physics hide inside the proton?*, Phys. Rev. Lett. **123** (2019) 13, 132001, arXiv:1905.05215 [hep-ph].

[143] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, JHEP **07** (2021) 122, arXiv:2104.02723 [hep-ph].

[144] A. M. Galda, M. Neubert, and S. Renner, *ALP — SMEFT interference*, JHEP **06** (2021) 135, arXiv:2105.01078 [hep-ph].

[145] A. Biekötter, J. Fuentes-Martín, A. M. Galda, and M. Neubert, *A global analysis of axion-like particle interactions using SMEFT fits*, JHEP **09** (2023) 120, arXiv:2307.10372 [hep-ph].

[146] A. Lessa and V. Sanz, *Going beyond Top EFT*, JHEP **04** (2024) 107, arXiv:2312.00670 [hep-ph].

[147] D. E. Morrissey, T. Plehn, and T. M. P. Tait, *Physics searches at the LHC*, Phys. Rept. **515** (2012) 1, arXiv:0912.3259 [hep-ph].

[148] M. Czakon, Z. Kassabov, A. Mitov, R. Poncelet, and A. Popescu, *HighTEA: high energy theory event analyser*, J. Phys. G **51** (2024) 11, 115002, arXiv:2304.05993 [hep-ph].

[149] ATLAS, G. Aad *et al.*, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, Phys. Lett. B **810** (2020) 135797, arXiv:2006.13076 [hep-ex].

[150] ATLAS, G. Aad *et al.*, *Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 8, 737, arXiv:2103.12603 [hep-ex].

[151] ATLAS, G. Aad *et al.*, *Measurement of single top-quark production in the s-channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2023) 191, arXiv:2209.08990 [hep-ex].

[152] S. Kraml *et al.*, *Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results*, Eur. Phys. J. C **72** (2012) 1976, arXiv:1203.2489 [hep-ph].

[153] F. Boudjema *et al.*, *On the presentation of the LHC Higgs Results*, in *Workshop on Likelihoods for the LHC Searches*. 7, 2013. arXiv:1307.5865 [hep-ph].

[154] K. Cranmer, S. Kreiss, D. Lopez-Val, and T. Plehn, *Decoupling Theoretical Uncertainties from Measurements of the Higgs Boson*, Phys. Rev. D **91** (2015) 5, 054032, arXiv:1401.0080 [hep-ph].

[155] M. Feickert, L. Heinrich, and G. Stark, *Likelihood preservation and statistical reproduction of searches for new physics*, EPJ Web Conf. **245** (2020) 06017.

[156] LHC Reinterpretation Forum, W. Abdallah *et al.*, *Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2*, SciPost Phys. **9** (2020) 2, 022, arXiv:2003.07868 [hep-ph].

[157] K. Cranmer *et al.*, *Publishing statistical models: Getting the most out of particle physics experiments*, SciPost Phys. **12** (2022) 1, 037, arXiv:2109.04981 [hep-ph].

[158] ROOT, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, .

[159] L. Heinrich, M. Feickert, and G. Stark, *pyhf: v0.7.6*, . https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6.

[160] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, J. Open Source Softw. **6** (2021) 58, 2823.

[161] J. Y. Araz, *Spey: Smooth inference for reinterpretation studies*, SciPost Phys. **16** (2024) 1, 032, arXiv:2307.06996 [hep-ph].

[162] A. Buckley, M. Citron, S. Fichet, S. Kraml, W. Waltenberger, and N. Wardle, *The Simplified Likelihood Framework*, JHEP **04** (2019) 064, arXiv:1809.05548 [hep-ph].

[163] M. Mahdi Altakach, S. Kraml, A. Lessa, S. Narasimha, T. Pascal, and W. Waltenberger, *SModelS v2.3: Enabling global likelihood analyses*, SciPost Phys. **15** (2023) 5, 185, arXiv:2306.17676 [hep-ph].

[164] CMS, V. Khachatryan *et al.*, *Measurement of the t-tbar production cross section in the e-mu channel in proton-proton collisions at sqrt(s) = 7 and 8 TeV*, JHEP **08** (2016) 029, arXiv:1603.02303 [hep-ex].

[165] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv:1112.5675 [hep-ph].

[166] ATLAS, M. Aaboud *et al.*, *Measurement of the inclusive and fiducial $t\bar{t}$ production cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 487, arXiv:1712.06857 [hep-ex].

[167] CMS, A. Tumasyan *et al.*, *Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **104** (2021) 9, 092013, arXiv:2108.02803 [hep-ex].

[168] CMS, A. M. Sirunyan *et al.*, *Measurement of the* $\mathrm{t\bar{t}}$ *production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 5, 368, arXiv:1812.10505 [hep-ex].

[169] ATLAS, G. Aad *et al.*, *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 6, 528, arXiv:1910.08819 [hep-ex].

[170] ATLAS, G. Aad *et al.*, *Measurements of top-quark pair single- and double-differential cross-sections in the all-hadronic channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, JHEP **01** (2021) 033, arXiv:2006.09274 [hep-ex].

[171] CMS, A. Tumasyan *et al.*, *First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV*, JHEP **08** (2023) 204, arXiv:2303.10680 [hep-ex].

[172] CMS, V. Khachatryan *et al.*, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 11, 542, arXiv:1505.04480 [hep-ex].

[173] M. Czakon, D. Heymes, and A. Mitov, *fastNLO tables for NNLO top-quark pair differential distributions*, arXiv:1704.08551 [hep-ph].

[174] M. Czakon, D. Heymes, and A. Mitov, *High-precision differential predictions for top-quark pairs at the LHC*, Phys. Rev. Lett. **116** (2016) 8, 082003, arXiv:1511.00549 [hep-ph].

[175] M. Czakon, D. Heymes, and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, JHEP **04** (2017) 071, arXiv:1606.03350 [hep-ph].

[176] ATLAS, G. Aad *et al.*, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 10, 538, arXiv:1511.04716 [hep-ex].

[177] CMS, A. M. Sirunyan *et al.*, *Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*, JHEP **02** (2019) 149, arXiv:1811.06625 [hep-ex].

[178] ATLAS, G. Aad *et al.*, *Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Eur. Phys. J. C **79** (2019) 12, 1028, arXiv:1908.07305 [hep-ex]. [Erratum: Eur.Phys.J.C 80, 1092 (2020)].

[179] ATLAS, G. Aad *et al.*, *Differential $t\bar{t}$ cross-section measurements using boosted top quarks in the all-hadronic final state with 139 fb^{-1} of ATLAS data*, JHEP **04** (2023) 080, arXiv:2205.02817 [hep-ex].

[180] CMS, V. Khachatryan *et al.*, *Inclusive and differential measurements of the $t\bar{t}$ charge asymmetry in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **757** (2016) 154, arXiv:1507.03119 [hep-ex].

[181] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, *Top-quark charge asymmetry at the LHC and Tevatron through NNLO QCD and NLO EW*, Phys. Rev. D **98** (2018) 1, 014003, arXiv:1711.03945 [hep-ph].

[182] CMS, V. Khachatryan *et al.*, *Measurements of $t\bar{t}$ charge asymmetry using dilepton final states in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **760** (2016) 365, arXiv:1603.06221 [hep-ex].

[183] ATLAS, G. Aad *et al.*, *Measurement of the charge asymmetry in top-quark pair production in the lepton-plus-jets final state in pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 2, 87, arXiv:1509.02358 [hep-ex]. [Erratum: Eur.Phys.J.C 77, 564 (2017)].

[184] ATLAS, G. Aad *et al.*, *Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Rev. D **94** (2016) 3, 032006, arXiv:1604.05538 [hep-ex].

[185] CMS, A. Tumasyan *et al.*, *Measurement of the tt^{-} charge asymmetry in events with highly Lorentz-boosted top quarks in pp collisions at s=13 TeV*, Phys. Lett. B **846** (2023) 137703, arXiv:2208.02751 [hep-ex].

[186] ATLAS, G. Aad *et al.*, *Evidence for the charge asymmetry in $pp \to t\bar{t}$ production at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2023) 077, arXiv:2208.12095 [hep-ex].

[187] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel, and V. Theeuwes, *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*, Eur. Phys. J. C **79** (2019) 3, 249, arXiv:1812.08622 [hep-ph].

[188] CMS, A. M. Sirunyan *et al.*, *Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2020) 056, arXiv:1907.11270 [hep-ex].

[189] CMS, A. M. Sirunyan *et al.*, *Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at* $\sqrt{s} =$ *13 TeV*, JHEP **08** (2018) 011, arXiv:1711.02547 [hep-ex].

[190] ATLAS, M. Aaboud *et al.*, *Measurement of the* $t\bar{t}Z$ *and* $t\bar{t}W$ *cross sections in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 7, 072009, arXiv:1901.03584 [hep-ex].

[191] CMS, A. M. Sirunyan *et al.*, *Measurement of the semileptonic* $\mathrm{t\bar{t}} + \gamma$ *production cross section in pp collisions at* $\sqrt{s} = 8$ *TeV*, JHEP **10** (2017) 006, arXiv:1706.08128 [hep-ex].

[192] ATLAS, M. Aaboud *et al.*, *Measurement of the* $t\bar{t}\gamma$ *production cross section in proton-proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, JHEP **11** (2017) 086, arXiv:1706.03046 [hep-ex].

[193] ATLAS, G. Aad *et al.*, *Comprehensive measurements of t-channel single top-quark production cross sections at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Phys. Rev. D **90** (2014) 11, 112006, arXiv:1406.7844 [hep-ex].

[194] CMS, S. Chatrchyan *et al.*, *Measurement of the Single-Top-Quark t-Channel Cross Section in pp Collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **12** (2012) 035, arXiv:1209.4533 [hep-ex].

[195] ATLAS, M. Aaboud *et al.*, *Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, Eur. Phys. J. C **77** (2017) 8, 531, arXiv:1702.02859 [hep-ex].

[196] CMS, V. Khachatryan *et al.*, *Measurement of the t-channel single-top-quark production cross section and of the* $| V_{tb} |$ *CKM matrix element in pp collisions at* $\sqrt{s}$*= 8 TeV*, JHEP **06** (2014) 090, arXiv:1403.7366 [hep-ex].

[197] ATLAS, M. Aaboud *et al.*, *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector*, JHEP **04** (2017) 086, arXiv:1609.03920 [hep-ex].

[198] E. L. Berger, J. Gao, C. P. Yuan, and H. X. Zhu, *NNLO QCD Corrections to t-channel Single Top-Quark Production and Decay*, Phys. Rev. D **94** (2016) 7, 071501, arXiv:1606.08463 [hep-ph].

[199] CMS, A. M. Sirunyan *et al.*, *Cross section measurement of t-channel single top quark production in pp collisions at* $\sqrt{s} =$ *13 TeV*, Phys. Lett. B **772** (2017) 752, arXiv:1610.00678 [hep-ex].

[200] CMS, A. M. Sirunyan *et al.*, *Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **80** (2020) 5, 370, arXiv:1907.08330 [hep-ex].

[201] CMS, V. Khachatryan *et al.*, *Search for s channel single top quark production in pp collisions at* $\sqrt{s} = 7$ *and 8 TeV*, JHEP **09** (2016) 027, arXiv:1603.02555 [hep-ex].

[202] ATLAS, G. Aad *et al.*, *Evidence for single top-quark production in the s-channel in proton-proton collisions at* $\sqrt{s}$ *=8 TeV with the ATLAS detector using the Matrix Element Method*, Phys. Lett. B **756** (2016) 228, arXiv:1511.05980 [hep-ex].

[203] ATLAS, G. Aad *et al.*, *Evidence for the associated production of a W boson and a top quark in ATLAS at* $\sqrt{s} = 7$ *TeV*, Phys. Lett. B **716** (2012) 142, arXiv:1205.5764 [hep-ex].

[204] CMS, S. Chatrchyan *et al.*, *Evidence for Associated Production of a Single Top Quark and W Boson in pp Collisions at* $\sqrt{s} = 7$ *TeV*, Phys. Rev. Lett. **110** (2013) 022003, arXiv:1209.3489 [hep-ex].

[205] ATLAS, G. Aad *et al.*, *Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment*, JHEP **01** (2016) 064, arXiv:1510.03752 [hep-ex].

[206] ATLAS, G. Aad *et al.*, *Measurement of single top-quark production in association with a W boson in the single-lepton channel at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 8, 720, arXiv:2007.01554 [hep-ex].

[207] CMS, S. Chatrchyan *et al.*, *Observation of the associated production of a single top quark and a W boson in pp collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Rev. Lett. **112** (2014) 23, 231802, arXiv:1401.2942 [hep-ex].

[208] ATLAS, M. Aaboud *et al.*, *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at* $\sqrt{s} = 13$ *TeV with ATLAS*, JHEP **01** (2018) 063, arXiv:1612.07231 [hep-ex].

[209] CMS, A. M. Sirunyan *et al.*, *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **10** (2018) 117, arXiv:1805.07399 [hep-ex].

[210] CMS, A. Tumasyan *et al.*, *Observation of tW production in the single-lepton channel in pp collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **11** (2021) 111, arXiv:2109.01706 [hep-ex].

[211] ATLAS, M. Aaboud *et al.*, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, Phys. Lett. B **780** (2018) 557, arXiv:1710.03659 [hep-ex].

[212] ATLAS, G. Aad *et al.*, *Measurement of the W boson polarization in top quark decays with the ATLAS detector*, JHEP **06** (2012) 088, arXiv:1205.2484 [hep-ex].

[213] CMS, S. Chatrchyan *et al.*, *Measurement of the W-Boson Helicity in Top-Quark decays from* $t\bar{t}$ *Production in Lepton + Jets Events in pp Collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **10** (2013) 167, arXiv:1308.3879 [hep-ex].

[214] ATLAS, M. Aaboud *et al.*, *Measurement of the W boson polarisation in* $t\bar{t}$ *events from pp collisions at* $\sqrt{s} = 8$ *TeV in the lepton + jets channel with ATLAS*, Eur. Phys. J. C **77** (2017) 4, 264, arXiv:1612.02577 [hep-ex]. [Erratum: Eur.Phys.J.C 79, 19 (2019)].

[215] CMS, V. Khachatryan *et al.*, *Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at* $\sqrt{s} = 8$*TeV*, Phys. Lett. B **762** (2016) 512, arXiv:1605.09047 [hep-ex].

[216] ATLAS, G. Aad *et al.*, *Measurement of the polarisation of W bosons produced in top-quark decays using dilepton events at s=13 TeV with the ATLAS experiment*, Phys. Lett. B **843** (2023) 137829, arXiv:2209.14903 [hep-ex].

[217] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *The automation of next-to-leading order electroweak calculations*, JHEP **07** (2018) 185, arXiv:1804.10017 [hep-ph]. [Erratum: JHEP 11, 085 (2021)].

[218] 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 [hep-ph].

[219] NNPDF, R. D. Ball *et al.*, *The path to proton structure at 1% accuracy*, Eur. Phys. J. C **82** (2022) 5, 428, arXiv:2109.02653 [hep-ph].

[220] 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** (2015) 132, arXiv:1412.7420 [hep-ph].

[221] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, Comput. Phys. Commun. **185** (2014) 2250, arXiv:1310.1921 [hep-ph].

[222] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Automated one-loop computations in the standard model effective field theory*, Phys. Rev. D **103** (2021) 9, 096024, arXiv:2008.11743 [hep-ph].

[223] A. Ghosh, B. Nachman, T. Plehn, L. Shire, T. M. P. Tait, and D. Whiteson, *Statistical patterns of theory uncertainties*, SciPost Phys. Core **6** (2023) 045, arXiv:2210.15167 [hep-ph].

[224] S. Dawson, S. Homiller, and S. D. Lane, *Putting standard model EFT fits to work*, Phys. Rev. D **102** (2020) 5, 055012, arXiv:2007.01296 [hep-ph].

[225] S. Dawson, S. Homiller, and M. Sullivan, *Impact of dimension-eight SMEFT contributions: A case study*, Phys. Rev. D **104** (2021) 11, 115013, arXiv:2110.06929 [hep-ph].

[226] I. Brivio, S. Bruggisser, E. Geoffray, W. Killian, M. Krämer, M. Luchmann, T. Plehn, and B. Summ, *From models to SMEFT and back?*, SciPost Phys. **12** (2022) 1, 036, arXiv:2108.01094 [hep-ph].

[227] A. Hocker, H. Lacker, S. Laplace, and F. Le Diberder, *A New approach to a global fit of the CKM matrix*, Eur. Phys. J. C **21** (2001) 225, arXiv:hep-ph/0104062.

[228] ATLAS, G. Aad *et al.*, *Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods*, .

[229] K. Cranmer and A. Held, *Building and steering binned template fits with cabinetry*, EPJ Web Conf. **251** (2021) 03067.

[230] C. Abel *et al.*, *Measurement of the Permanent Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **124** (2020) 8, 081803, arXiv:2001.11966 [hep-ex].

[231] B. C. Regan, E. D. Commins, C. J. Schmidt, and D. DeMille, *New limit on the electron electric dipole moment*, Phys. Rev. Lett. **88** (2002) 071805.

[232] S. A. Murthy, D. Krause, Z. L. Li, and L. R. Hunter, *New Limits on the Electron Electric Dipole Moment from Cesium*, Phys. Rev. Lett. **63** (1989) 965.

[233] T. S. Roussy *et al.*, *An improved bound on the electron's electric dipole moment*, Science **381** (2023) 6653, adg4084, arXiv:2212.11841 [physics.atom-ph].

[234] ACME, V. Andreev *et al.*, *Improved limit on the electric dipole moment of the electron*, Nature **562** (2018) 7727, 355.

[235] J. J. Hudson, D. M. Kara, I. J. Smallman, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, *Improved measurement of the shape of the electron*, Nature **473** (2011) 493.

[236] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, *Reduced Limit on the Permanent Electric Dipole Moment of Hg199*, Phys. Rev. Lett. **116** (2016) 16, 161601, arXiv:1601.04339 [physics.atom-ph]. [Erratum: Phys.Rev.Lett. 119, 119901 (2017)].

[237] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, *Erratum: Reduced limit on the permanent electric dipole moment of* ^{199}Hg *[phys. rev. lett. 116, 161601 (2016)]*, Phys. Rev. Lett. **119** (Sep, 2017) 119901.

[238] N. Sachdeva *et al.*, *New Limit on the Permanent Electric Dipole Moment of* ^{129}Xe *using* ^{3}He *Comagnetometry and SQUID Detection*, Phys. Rev. Lett. **123** (2019) 14, 143003, arXiv:1902.02864 [physics.atom-ph].

[239] F. Allmendinger, I. Engin, W. Heil, S. Karpuk, H.-J. Krause, B. Niederländer, A. Offenhäusser, M. Repetto, U. Schmidt, and S. Zimmer, *Measurement of the permanent electric dipole moment of the* ^{129}Xe *atom*, Phys. Rev. A **100** (Aug, 2019) 022505.

[240] T. A. Zheng, Y. A. Yang, S. Z. Wang, J. T. Singh, Z. X. Xiong, T. Xia, and Z. T. Lu, *Measurement of the Electric Dipole Moment of Yb171 Atoms in an Optical Dipole Trap*, Phys. Rev. Lett. **129** (2022) 8, 083001, arXiv:2207.08140 [physics.atom-ph].

[241] M. Bishof *et al.*, *Improved limit on the* ^{225}Ra *electric dipole moment*, Phys. Rev. C **94** (2016) 2, 025501, arXiv:1606.04931 [nucl-ex].

[242] D. Cho, K. Sangster, and E. A. Hinds, *Search for time reversal symmetry violation in thallium fluoride using a jet source*, Phys. Rev. A **44** (1991) 2783.

[243] E. S. Ensberg, *Experimental upper limit for the permanent electric dipole moment of* rb^{85} *by optical-pumping techniques*, Phys. Rev. **153** (Jan, 1967) 36.

[244] M. A. Player and P. G. H. Sandars, *An Experiment to Search for an Electric Dipole Moment in the P(2)-3 Metastable State of Xenon*, J. Phys. B **3** (1970) 1620.

[245] S. Eckel, P. Hamilton, E. Kirilov, H. W. Smith, and D. DeMille, *Search for the electron electric dipole moment using* Ω*-doublet levels in PbO*, Phys. Rev. A **87** (2013) 5, 052130, arXiv:1303.3075 [physics.atom-ph].

[246] S. Eckel, A. O. Sushkov, and S. K. Lamoreaux, *A limit on the electron electric dipole moment using paramagnetic ferroelectric* $Eu_{0.5}Ba_{0.5}TiO_3$, Phys. Rev. Lett. **109** (2012) 193003, arXiv:1208.4420 [physics.atom-ph].

[247] L. Pondrom, R. Handler, M. Sheaff, P. T. Cox, J. Dworkin, O. E. Overseth, T. Devlin, L. Schachinger, and K. J. Heller, *New Limit on the Electric Dipole Moment of the* Λ *Hyperon*, Phys. Rev. D **23** (1981) 814.

[248] P. V. Coveney and P. G. H. Sandars, *PARITY AND TIME VIOLATING INTERACTIONS IN THALLIUM FLUORIDE*, J. Phys. B **16** (1983) 3727.

[249] T. Fleig and L. V. Skripnikov, *$\mathcal{P},\mathcal{T}$-Violating and Magnetic Hyperfine Interactions in Atomic Thallium*, Symmetry **12** (2020) 4, 498, arXiv:1910.11596 [physics.atom-ph].

[250] J. S. M. Ginges and V. V. Flambaum, *Violations of fundamental symmetries in atoms and tests of unification theories of elementary particles*, Phys. Rept. **397** (2004) 63, arXiv:physics/0309054.

[251] T. Fleig and M. Jung, *Model-independent determinations of the electron EDM and the role of diamagnetic atoms*, JHEP **07** (2018) 012, arXiv:1802.02171 [hep-ph].

[252] V. A. Dzuba, V. V. Flambaum, and S. G. Porsev, *Calculation of P,T-odd electric dipole moments for diamagnetic atoms Xe-129, Yb-171, Hg-199, Rn-211, and Ra-225*, Phys. Rev. A **80** (2009) 032120, arXiv:0906.5437 [physics.atom-ph].

[253] K. Gaul and R. Berger, *Global analysis of $\mathcal{CP}$-violation in atoms, molecules and role of medium-heavy systems*, JHEP **08** (2024) 100, arXiv:2312.08858 [hep-ph].

[254] T. Fleig, *P,T-odd weak neutral current interactions in the TlF molecule*, Phys. Rev. A **109** (2024) 2, 022807, arXiv:2311.06376 [physics.atom-ph].

[255] J. de Vries, P. Draper, K. Fuyuto, J. Kozaczuk, and B. Lillard, *Uncovering an axion mechanism with the EDM portfolio*, Phys. Rev. D **104** (2021) 5, 055039, arXiv:2107.04046 [hep-ph].

[256] Particle Data Group, R. L. Workman *et al.*, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[257] A.-M. Mårtensson-Pendrill and P. Öster, *Calculations of atomic electric dipole moments*, Physica Scripta **36** (1987) 3, 444.

[258] L. V. Skripnikov, *Communication: Theoretical study of hff+ cation to search for the t,p-odd interactions*, The Journal of Chemical Physics **147** (jul, 2017) .

[259] T. Fleig, *$\mathcal{P},\mathcal{T}$-odd and magnetic hyperfine-interaction constants and excited-state lifetime for HfF^+*, Phys. Rev. A **96** (2017) 4, 040502, arXiv:1706.02893 [physics.atom-ph].

[260] A. N. Petrov, N. S. Mosyagin, T. A. Isaev, and A. V. Titov, *Theoretical study of Hf F+ in search of the electron electric dipole moment*, Phys. Rev. A **76** (2007) 030501, arXiv:physics/0611254.

[261] E. R. Meyer, J. L. Bohn, and M. P. Deskevich, *Candidate molecular ions for an electron electric dipole moment experiment*, Phys. Rev. A **73** (Jun, 2006) 062108.

[262] L. V. Skripnikov, *Combined 4-component and relativistic pseudopotential study of ThO for the electron electric dipole moment search*, The Journal of Chemical Physics **145** (12, 2016) 214301, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4968229/15523128/214301_1_online.pdf.

[263] E. R. Meyer and J. L. Bohn, *Prospects for an electron electric dipole moment search in metastable ThO and ThF*, Phys. Rev. A **78** (2008) 010502, arXiv:0805.0161 [physics.atom-ph].

[264] M. Denis and T. Fleig, *In search of discrete symmetry violations beyond the standard model: Thorium monoxide reloaded*, The Journal of Chemical Physics **145** (12, 2016) 214307, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4968597/15523220/214307_1_online.pdf.

[265] V. A. Dzuba, V. V. Flambaum, and C. Harabati, *Relations between matrix elements of different weak interactions and interpretation of the parity-nonconserving and electron electric-dipole-moment measurements in atoms and molecules*, Phys. Rev. A **84** (Nov, 2011) 052108.

[266] V. A. Dzuba, V. V. Flambaum, and C. Harabati, *Erratum: Relations between matrix elements of different weak interactions and interpretation of the parity-nonconserving and electron electric-dipole-moment measurements in atoms and molecules [phys. rev. a 84, 052108 (2011)]*, Phys. Rev. A **85** (Feb, 2012) 029901.

[267] A. Sunaga, M. Abe, M. Hada, and B. P. Das, *Relativistic coupled-cluster calculation of the electron-nucleus scalar-pseudoscalar interaction constant W_s in ybf*, Phys. Rev. A **93** (2016) 042507.

[268] M. Abe, G. Gopakumar, M. Hada, B. P. Das, H. Tatewaki, and D. Mukherjee, *Application of relativistic coupled-cluster theory to the effective electric field in ybf*, Physical Review A **90** (2014) 2, .

[269] Y. Ema, T. Gao, and M. Pospelov, *Improved Indirect Limits on Muon Electric Dipole Moment*, Phys. Rev. Lett. **128** (2022) 13, 131803, arXiv:2108.05398 [hep-ph].

[270] ACME, J. Baron *et al.*, *Methods, Analysis, and the Treatment of Systematic Errors for the Electron Electric Dipole Moment Search in Thorium Monoxide*, New J. Phys. **19** (2017) 7, 073029, arXiv:1612.09318 [physics.atom-ph].

[271] T. Fleig and M. Jung, *$\mathcal{P},\mathcal{T}$ -odd interactions in atomic ^{129}Xe and phenomenological applications*, Phys. Rev. A **103** (2021) 1, 012807, arXiv:2009.07730 [hep-ph].

[272] V. Flambaum and I. Khriplovich, *New bounds on the electric dipole moment of the electron and on t-odd electron-nucleon coupling*, Zh. Eksp. Theor. Fiz **89** (1985) 1505.

[273] K. Yanase and N. Shimizu, *Large-scale shell-model calculations of nuclear Schiff moments of ^{129}Xe and ^{199}Hg*, Phys. Rev. C **102** (2020) 6, 065502, arXiv:2006.15142 [nucl-th].

[274] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, *Activation functions: Comparison of trends in practice and research for deep learning*, arXiv:1811.03378 [cs.LG].

[275] D. Hendrycks and K. Gimpel, *Gaussian error linear units (gelus)*, arXiv:1606.08415 [cs.LG].

[276] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980 [cs.LG].

[277] I. Loshchilov and F. Hutter, *Decoupled weight decay regularization*, arXiv:1711.05101 [cs.LG].

[278] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, *Variational inference: A review for statisticians*, Journal of the American Statistical Association **112** (Apr., 2017) 859–877.

[279] G. E. Hinton and D. van Camp, *Keeping the neural networks simple by minimizing the description length of the weights*, in *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, COLT '93. Association for Computing Machinery, New York, NY, USA, 1993.

[280] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, The Annals of Mathematical Statistics **22** (1951) 1, 79 .

[281] D. P. Kingma, T. Salimans, and M. Welling, *Variational dropout and the local reparameterization trick*, in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds. Curran Associates, Inc., 2015.

[282] T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause, and R. Winterhalder, *Modern Machine Learning for LHC Physicists*, arXiv:2211.01421 [hep-ph].

[283] A. Power, Y. Burda, H. Edwards, I. Babuschkin, and V. Misra, *Grokking: Generalization beyond overfitting on small algorithmic datasets*, CoRR **abs/2201.02177** (2022) , 2201.02177.

[284] A. Butter, N. Huetsch, S. Palacios Schweitzer, T. Plehn, P. Sorrenson, and J. Spinner, *Jet Diffusion versus JetGPT – Modern Networks for the LHC*, arXiv:2305.10475 [hep-ph].

[285] Q. Liu and D. Wang, *Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm*, arXiv:1608.04471 [stat.ML].

[286] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, *KAN: Kolmogorov-Arnold Networks*, arXiv:2404.19756 [cs.LG].

[287] Z. Liu, P. Ma, Y. Wang, W. Matusik, and M. Tegmark, *KAN 2.0: Kolmogorov-Arnold Networks Meet Science*, arXiv:2408.10205 [cs.LG].

[288] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, JHEP **01** (2019) 121, arXiv:1810.05165 [hep-ph].

[289] F. Girosi and T. Poggio, *Representation properties of networks: Kolmogorov's theorem is irrelevant*, Neural Computation **1** (1989) 4, 465.

[290] A. Molina, P. Schramowski, and K. Kersting, *Padé Activation Units: End-to-end Learning of Flexible Activation Functions in Deep Networks*, arXiv:1907.06732 [cs.LG].

[291] X. Yang and X. Wang, *Kolmogorov-Arnold Transformer*, arXiv:2409.10594 [cs.LG].

[292] J. Erdmann, F. Mausolf, and J. L. Späh, *KAN we improve on HEP classification tasks? Kolmogorov-Arnold Networks applied to an LHC physics example*, arXiv:2408.02743 [hep-ph].

[293] E. Abasov, P. Volkov, G. Vorotnikov, L. Dudko, A. Zaborenko, E. Iudin, A. Markina, and M. Perfilov, *Application of Kolmogorov-Arnold Networks in high energy physics*, arXiv:2409.01724 [hep-ph].

[294] T. Heimel, R. Winterhalder, A. Butter, J. Isaacson, C. Krause, F. Maltoni, O. Mattelaer, and T. Plehn, *MadNIS - Neural multi-channel importance sampling*, SciPost Phys. **15** (2023) 4, 141, arXiv:2212.06172 [hep-ph].

[295] T. Heimel, N. Huetsch, F. Maltoni, O. Mattelaer, T. Plehn, and R. Winterhalder, *The MadNIS reloaded*, SciPost Phys. **17** (2024) 1, 023, arXiv:2311.01548 [hep-ph].

[296] N. Deutschmann and N. Götz, *Accelerating HEP simulations with Neural Importance Sampling*, JHEP **03** (2024) 083, arXiv:2401.09069 [hep-ph].

[297] T. Heimel, O. Mattelaer, T. Plehn, and R. Winterhalder, *Differentiable MadNIS-Lite*, arXiv:2408.01486 [hep-ph].

[298] D. Maître and H. Truong, *One-loop matrix element emulation with factorisation awareness*, arXiv:2302.04005 [hep-ph].

[299] J. Spinner, V. Bresó, P. de Haan, T. Plehn, J. Thaler, and J. Brehmer, *Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics*, arXiv:2405.14806 [physics.data-an].

[300] D. Maître, V. S. Ngairangbam, and M. Spannowsky, *Optimal Equivariant Architectures from the Symmetries of Matrix-Element Likelihoods*, arXiv:2410.18553 [hep-ph].

[301] J. Brehmer, V. Bresó, P. de Haan, T. Plehn, H. Qu, J. Spinner, and J. Thaler, *A Lorentz-Equivariant Transformer for All of the LHC*, arXiv:2411.00446 [hep-ph].

[302] R. Winterhalder, V. Magerya, E. Villa, S. P. Jones, M. Kerner, A. Butter, G. Heinrich, and T. Plehn, *Targeting multi-loop integrals with neural networks*, SciPost Phys. **12** (2022) 4, 129, arXiv:2112.09145 [hep-ph].

[303] B. Hashemi, N. Amin, K. Datta, D. Olivito, and M. Pierini, *LHC analysis-specific datasets with Generative Adversarial Networks*, arXiv:1901.05282 [hep-ex].

[304] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, and S. Palazzo, *DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC*, JHEP **08** (2019) 110, arXiv:1903.02433 [hep-ex].

[305] A. Butter, T. Plehn, and R. Winterhalder, *How to GAN LHC Events*, SciPost Phys. **7** (2019) 6, 075, arXiv:1907.03764 [hep-ph].

[306] Y. Alanazi, N. Sato, T. Liu, W. Melnitchouk, M. P. Kuchera, E. Pritchard, M. Robertson, R. Strauss, L. Velasco, and Y. Li, *Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)*, arXiv:2001.11103 [hep-ph].

[307] M. Paganini, L. de Oliveira, and B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, Phys. Rev. Lett. **120** (2018) 4, 042003, arXiv:1705.02355 [hep-ex].

[308] M. Paganini, L. de Oliveira, and B. Nachman, *CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, Phys. Rev. D **97** (2018) 1, 014021, arXiv:1712.10321 [hep-ex].

[309] M. Erdmann, J. Glombitza, and T. Quast, *Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network*, Comput. Softw. Big Sci. **3** (2019) 1, 4, arXiv:1807.01954 [physics.ins-det].

[310] D. Belayneh *et al.*, *Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics*, Eur. Phys. J. C **80** (12, 2020) 688, arXiv:1912.06794 [physics.ins-det].

[311] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, and K. Krüger, *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*, Comput. Softw. Big Sci. **5** (2021) 1, 13, arXiv:2005.05334 [physics.ins-det].

[312] C. Krause and D. Shih, *Fast and accurate simulations of calorimeter showers with normalizing flows*, Phys. Rev. D **107** (2023) 11, 113003, arXiv:2106.05285 [physics.ins-det].

[313] ATLAS Collaboration, *AtlFast3: the next generation of fast simulation in ATLAS*, Comput. Softw. Big Sci. **6** (2022) 7, arXiv:2109.02551 [hep-ex].

[314] C. Krause and D. Shih, *Accelerating accurate simulations of calorimeter showers with normalizing flows and probability density distillation*, Phys. Rev. D **107** (2023) 11, 113004, arXiv:2110.11377 [physics.ins-det].

[315] E. Buhmann, S. Diefenbacher, D. Hundhausen, G. Kasieczka, W. Korcari, E. Eren, F. Gaede, K. Krüger, P. McKeown, and L. Rustige, *Hadrons, better, faster, stronger*, Mach. Learn. Sci. Tech. **3** (2022) 2, 025014, arXiv:2112.09709 [physics.ins-det].

[316] C. Chen, O. Cerri, T. Q. Nguyen, J. R. Vlimant, and M. Pierini, *Analysis-Specific Fast Simulation at the LHC with Deep Learning*, Comput. Softw. Big Sci. **5** (2021) 1, 15.

[317] V. Mikuni and B. Nachman, *Score-based generative models for calorimeter shower simulation*, Phys. Rev. D **106** (2022) 9, 092009, arXiv:2206.11898 [hep-ph].

[318] J. C. Cresswell, B. L. Ross, G. Loaiza-Ganem, H. Reyes-Gonzalez, M. Letizia, and A. L. Caterini, *CaloMan: Fast generation of calorimeter showers with density estimation on learned manifolds*, in *36th Conference on Neural Information Processing Systems*. 11, 2022. arXiv:2211.15380 [hep-ph].

[319] S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, C. Krause, I. Shekhzadeh, and D. Shih, *L2LFlows: generating high-fidelity 3D calorimeter images*, JINST **18** (2023) 10, P10017, arXiv:2302.11594 [physics.ins-det].

[320] A. Xu, S. Han, X. Ju, and H. Wang, *Generative machine learning for detector response modeling with a conditional normalizing flow*, JINST **19** (2024) 02, P02003, arXiv:2303.10148 [hep-ex].

[321] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, W. Korcari, K. Krüger, and P. McKeown, *CaloClouds: fast geometry-independent highly-granular calorimeter simulation*, JINST **18** (2023) 11, P11025, arXiv:2305.04847 [physics.ins-det].

[322] M. R. Buckley, C. Krause, I. Pang, and D. Shih, *Inductive simulation of calorimeter showers with normalizing flows*, Phys. Rev. D **109** (2024) 3, 033006, arXiv:2305.11934 [physics.ins-det].

[323] S. Diefenbacher, V. Mikuni, and B. Nachman, *Refining Fast Calorimeter Simulations with a Schrödinger Bridge*, arXiv:2308.12339 [physics.ins-det].

[324] F. Ernst, L. Favaro, C. Krause, T. Plehn, and D. Shih, *Normalizing Flows for High-Dimensional Detector Simulations*, arXiv:2312.09290 [hep-ph].

[325] L. Favaro, A. Ore, S. P. Schweitzer, and T. Plehn, *CaloDREAM – Detector Response Emulation via Attentive flow Matching*, arXiv:2405.09629 [hep-ph].

[326] T. Buss, F. Gaede, G. Kasieczka, C. Krause, and D. Shih, *Convolutional L2LFlows: Generating Accurate Showers in Highly Granular Calorimeters Using Convolutional Normalizing Flows*, arXiv:2405.20407 [physics.ins-det].

[327] G. Quétant, J. A. Raine, M. Leigh, D. Sengupta, and T. Golling, *PIPPIN: Generating variable length full events from partons*, arXiv:2406.13074 [hep-ph].

[328] O. Amram *et al.*, *CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation*, arXiv:2410.21611 [cs.LG].

[329] A. Butter, S. Diefenbacher, G. Kasieczka, B. Nachman, and T. Plehn, *GANplifying event samples*, SciPost Phys. **10** (2021) 6, 139, arXiv:2008.06545 [hep-ph].

[330] S. Bieringer, A. Butter, S. Diefenbacher, E. Eren, F. Gaede, D. Hundhausen, G. Kasieczka, B. Nachman, T. Plehn, and M. Trabs, *Calomplification — the power of generative calorimeter models*, JINST **17** (2022) 09, P09028, arXiv:2202.07352 [hep-ph].

[331] A. Butter, T. Heimel, S. Hummerich, T. Krebs, T. Plehn, A. Rousselot, and S. Vent, *Generative networks for precision enthusiasts*, SciPost Phys. **14** (2023) 4, 078, arXiv:2110.13632 [hep-ph].

[332] R. Winterhalder, M. Bellagente, and B. Nachman, *Latent Space Refinement for Deep Generative Models*, arXiv:2106.00792 [stat.ML].

[333] B. Nachman and R. Winterhalder, *Elsa: enhanced latent spaces for improved collider simulations*, Eur. Phys. J. C **83** (2023) 9, 843, arXiv:2305.07696 [hep-ph].

[334] M. Leigh, D. Sengupta, J. A. Raine, G. Quétant, and T. Golling, *Faster diffusion model with improved quality for particle cloud generation*, Phys. Rev. D **109** (2024) 1, 012010, arXiv:2307.06836 [hep-ex].

[335] R. Das, L. Favaro, T. Heimel, C. Krause, T. Plehn, and D. Shih, *How to understand limitations of generative networks*, SciPost Phys. **16** (2024) 1, 031, arXiv:2305.16774 [hep-ph].

[336] S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn, and J. Thompson, *Deep-Learning Jets with Uncertainties and More*, SciPost Phys. **8** (2020) 1, 006, arXiv:1904.10004 [hep-ph].

[337] G. Kasieczka, M. Luchmann, F. Otterpohl, and T. Plehn, *Per-Object Systematics using Deep-Learned Calibration*, SciPost Phys. **9** (2020) 089, arXiv:2003.11099 [hep-ph].

[338] R. Gambhir, B. Nachman, and J. Thaler, *Learning Uncertainties the Frequentist Way: Calibration and Correlation in High Energy Physics*, Phys. Rev. Lett. **129** (2022) 8, 082001, arXiv:2205.03413 [hep-ph].

[339] H. Du, C. Krause, V. Mikuni, B. Nachman, I. Pang, and D. Shih, *Unifying Simulation and Inference with Normalizing Flows*, arXiv:2404.18992 [hep-ph].

[340] ATLAS, G. Aad *et al.*, *Precision calibration of calorimeter signals in the ATLAS experiment using an uncertainty-aware neural network*, arXiv:2412.04370 [hep-ex].

[341] M. Bellagente, M. Haussmann, M. Luchmann, and T. Plehn, *Understanding Event-Generation Networks via Uncertainties*, SciPost Phys. **13** (2022) 1, 003, arXiv:2104.04543 [hep-ph].

[342] S. Bieringer, S. Diefenbacher, G. Kasieczka, and M. Trabs, *Calibrating Bayesian generative machine learning for Bayesiamplification*, Mach. Learn. Sci. Tech. **5** (2024) 4, 045044, arXiv:2408.00838 [cs.LG].

[343] Sherpa, E. Bothmann *et al.*, *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 3, 034, arXiv:1905.09127 [hep-ph].

[344] S. Badger, B. Biedermann, P. Uwer, and V. Yundin, *Numerical evaluation of virtual corrections to multi-jet production in massless QCD*, Comput. Phys. Commun. **184** (2013) 1981, arXiv:1209.0100 [hep-ph].

[345] M. Hubert and T. Fleig, *Electric dipole moments generated by nuclear Schiff moment interactions: A reassessment of the atoms Xe129 and Hg199 and the molecule TlF205*, Phys. Rev. A **106** (2022) 2, 022817, arXiv:2203.04618 [physics.atom-ph].

[346] V. V. Flambaum and V. A. Dzuba, *Electric dipole moments of atoms and molecules produced by enhanced nuclear Schiff moments*, Phys. Rev. A **101** (2020) 4, 042504, arXiv:1912.03598 [physics.atom-ph].

[347] T. Chupp and M. Ramsey-Musolf, *Electric Dipole Moments: A Global Analysis*, Phys. Rev. C **91** (2015) 3, 035502, arXiv:1407.1064 [hep-ph].

[348] V. V. Flambaum, I. B. Khriplovich, and O. P. Sushkov, *On the P and T Nonconserving Nuclear Moments*, Nucl. Phys. A **449** (1986) 750.

[349] V. F. Dmitriev, R. A. Sen'kov, and N. Auerbach, *Effects of core polarization on the nuclear Schiff moment*, Phys. Rev. C **71** (2005) 035501, arXiv:nucl-th/0408065.

[350] V. A. Dzuba, V. V. Flambaum, and J. S. M. Ginges, *Atomic electric dipole moments of He and Yb induced by nuclear Schiff moments*, Phys. Rev. A **76** (2007) 034501, arXiv:0705.4001 [physics.atom-ph].

[351] J. Dobaczewski and J. Engel, *Nuclear time-reversal violation and the Schiff moment of Ra-225*, Phys. Rev. Lett. **94** (2005) 232502, arXiv:nucl-th/0503057.

[352] V. A. Dzuba, V. V. Flambaum, and P. G. Silvestrov, *Bounds on Electric Dipole Moments and T Violating Weak Interactions of the Nucleons*, Phys. Lett. B **154** (1985) 93.

[353] I. B. Khriplovich, *Limit on the Proton Electric Dipole Moment from Atomic Experiments*, Sov. Phys. JETP **44** (1976) 25.

[354] I. B. Khriplovich and S. K. Lamoreaux, *CP violation without strangeness: Electric dipole moments of particles, atoms, and molecules*. Springer Berlin Heidelberg, 1997.

[355] V. V. Flambaum, D. DeMille, and M. G. Kozlov, *Time-reversal symmetry violation in molecules induced by nuclear magnetic quadrupole moments*, Phys. Rev. Lett. **113** (2014) 103003, arXiv:1406.6479 [physics.atom-ph].