This Dissertation

entitled

SEARCH FOR NEW PHYSICS IMPACTING ASSOCIATED TOP PRODUCTION IN MULTILEPTON FINAL STATES USING THE FRAMEWORK OF EFFECTIVE FIELD THEORY

typeset with NDdiss2$_\varepsilon$ v3.2017.2 (2017/05/09) on October 23, 2022 for

Kelci Mohrman

This LaTeX 2$_\varepsilon$ classfile conforms to the University of Notre Dame style guidelines as of Fall 2012. However it is still possible to generate a non-conformant document if the instructions in the class file documentation are not followed!

> Be sure to refer to the published Graduate School guidelines at `http://graduateschool.nd.edu` as well. Those guidelines override everything mentioned about formatting in the documentation for this NDdiss2$_\varepsilon$ class file.

This page can be disabled by specifying the "`noinfo`*" option to the class invocation.* (i.e.,`\documentclass[...,noinfo]{nddiss2e}`)

This page is *NOT* part of the dissertation/thesis. It should be disabled before making final, formal submission, but should be included in the version submitted for format check.

NDdiss2$_\varepsilon$ documentation can be found at these locations:

`http://graduateschool.nd.edu`
`https://ctan.org/pkg/nddiss`

SEARCH FOR NEW PHYSICS IMPACTING ASSOCIATED TOP PRODUCTION IN MULTILEPTON FINAL STATES USING THE FRAMEWORK OF EFFECTIVE FIELD THEORY

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Kelci Mohrman

Kevin Lannon, Director

Graduate Program in Physics
Notre Dame, Indiana

October 2022

SEARCH FOR NEW PHYSICS IMPACTING ASSOCIATED TOP PRODUCTION IN MULTILEPTON FINAL STATES USING THE FRAMEWORK OF EFFECTIVE FIELD THEORY

Abstract

by

Kelci Mohrman

The abstract.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

CHAPTER 1

INTRODUCTION

The goal of this analysis is to search for new physics impacting associated top production in multi-lepton final states, using the framework of effective field theory (EFT) to parametrize the potential new physics effects.

While there are many compelling indications that the standard model (SM) of particle physics does not provide a complete description of nature (e.g. the strong evidence for dark matter [7, 16] and dark energy [15]), there is no a priori reason to assume new particles must exist in the energy range that is directly accessible at the LHC. If new physics particles are too heavy to be produced on-shell at the LHC, it may be difficult to identify their signatures with a direct search. However, an approach that indirectly probes higher energy scales may be able to discover these particles via their off-shell effects. The center of mass energy for collisions at the LCH will not significantly increase throughout the LHC's remaining years of operation, so indirect approaches may provide an exciting opportunity to extend the discovery reach of the LHC. EFT is an example of such an indirect probe; as a flexible method of systematically describing the off-shell effects of heavy new particles, EFT represents an important part of the search for new physics at the energy frontier.

In general, an effective theory is an approximation, valid under a certain energy range, for a more fundamental underlying theory. In SM effective field theory (SM EFT), the SM is treated as the lowest order term in an expansion of higher dimensional operators; the operators are constructed from products of SM fields that obey the symmetries of the SM. The EFT operators describe new physics interactions at a mass scale Λ. The strengths of the new physics interactions are described by dimensionless parameters known as Wilson Coefficients (WCs). The EFT Lagrangian can thus be expressed as follows:

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda}\mathcal{O}_i^5 + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^6 + ..., \tag{1.1}$$

where $\mathcal{L}_{\mathrm{SM}}$ is the SM Lagrangian, $\mathcal{O}_i^d$ are the EFT operators of dimension d, and c_i^d are the WCs for the operators of dimension d. Since each order in the expansion is scaled by an additional power of Λ, the terms in the lowest orders are expected to contribute the most significantly. This analysis therefore focuses on dimension-six operators, as these are the lowest order terms that contribute. The EFT framework will be discussed in more detail in Section 2.2.

While many analyses target a specific signature predicted by a particular new physics model, the EFT approach is more general. Assuming that the SM Lagrangian is the correct and complete description of all physics that is light enough to be probed directly with current experimental capabilities, the EFT Lagrangian provides a systematic description of the off-shell effects of heavy new physics scenarios, allowing for a consistent method of describing these effects across multiple sectors. EFT is thus a complementary approach to dedicated searches; if the off-shell effects of new physics manifest in a variety of signatures across many final states, a global EFT approach may be capable of identifying a statistically significant observation of the combination of effects, even if the effects are not significant when studied individually.

Although the ultimate goal of the EFT paradigm would be a global combination across all sectors of study at the LHC, the first step towards this goal is to begin performing EFT analyses within individual sectors. The analysis described in this thesis focuses on the top sector, targeting processes in which top quarks are produced in association with additional charged leptons. In the SM, these signatures are primarily produced by $t(\bar{t})X$ processes, where the X is a H, W, or Z boson. We refer to these processes as associated top production. Involving multiple heavy particles, the processes are relatively rare, and we are just now reaching the point where we have accumulated enough statistics to study these processes in detail; for these reasons, associated top processes may be an interesting venue in which to stage a search for new physics. In accordance with the global mindset of the EFT approach, this analysis aims to study all dimension-six EFT operators (involving top quarks) that can significantly impact associated top production processes.

The full set of associated top processes studied in this analysis is $t\bar{t}H$, $t\bar{t}l\nu$, $t\bar{t}l\bar{l}$, $tl\bar{l}q$, tHq, and $t\bar{t}t\bar{t}$. These processes give rise to a variety of final state signatures; in this analysis, we choose to focus on signatures involving multiple charged leptons. Referred to as multilepton final states, these signatures contain 2 leptons of the same charge or contain 3 or more leptons. Multilepton final states have relatively few backgrounds, clean detector signatures, and efficient triggers. In spite of these experimental benefits, a multilepton EFT analysis also gives rise to several challenges. These challenges primarily stem from the fact that many different processes and effects are capable of contributing to the same final state multilepton signatures. For example, if we consider a final state with two leptons of the same charge, we would expect contributions from both SM $t\bar{t}W$ and SM $t\bar{t}H$ production (as well as a contribution from SM $t\bar{t}Z$ when one of the leptons is lost). Many different dimension-six EFT operators can impact to these processes, interfering with each other and with the SM, making this final state a complicated admixture of processes and effects. Other multilepton final states will contain similarly complex admixtures of processes and EFT effects.

Because these effects cannot be isolated from each other, it is important to analyze the effects of all relevant operators across all channels simultaneously. For this reason, it would be difficult to construct a multi lepton EFT analysis as a reinterpretation of inclusive or differential cross section measurements. Instead, we make use of an approach that directly targets the EFT effects at detector level. First developed in Ref. [18], the key idea of this approach is the parameterization of the predicted yields in terms of the WCs. The procedure through which we obtain this parametrization will be detailed in Section 3.2.2.

Making use of more than three times as much data as was available for [18], the analysis described in this thesis builds on the techniques and tools developed in [18], improving on [18] in several key ways. Since [18] was performed with limited statistics, only inclusive categories defined by the multiplicity of final state objects were studied; with the increased statistics, this analysis leverages differential kinematic distributions within each inclusive bin, allowing additional sensitivity to be gained. An additional signal process ($\mathrm{t\bar{t}t\bar{t}}$) and 10 more dimension-six EFT operators are also included, brining the total number of WCs to 26. These improvements allow stronger limits to be placed on the WCs, resulting in a better understanding of the possibility of heavy new physics effects in the top sector.

The chapters of this thesis are organized in the following order. In Chapter 2, the theoretical concepts of the SM and of the EFT framework are discussed. Chapter 3 describes the simulated samples used in the analysis (including a discussion of the EFT parametrization of the signal samples). In Chapter 4, the CMS detector is described, and Chapter 5 explains how the particle reconstruction is performed. The event selection is detailed in Chapter 6. Chapter 7 describes the data to Monte Carlo corrections for simulated events. The backgrounds for this analysis are discussed in Chapter 8. Chapter 9 enumerates the systematic uncertainties of the analysis. The statistical tools used to extract the confidence intervals for the WCs are explained in Chapter 10. Chapter 11 presents the results of the analysis. A summary is provided in Chapter 12.

CHAPTER 2

THEORY

This chapter will discuss the theoretical background for this analysis. The SM is summarized in Section 2.1, and the relevant aspects of the SM EFT framework are discussed in Section 2.2.

2.1 The standard model

The standard model (SM) of particle physics is the mathematical framework that describes fundamental particles and their interactions. The SM is a quantum field theory with $\mathrm{SU}(3)_c \times \mathrm{SU}(2) \times \mathrm{U}(1)$ gauge symmetry. The $\mathrm{SU}(3)_c$ component corresponds to QCD, and is non-abelian. The $\mathrm{SU}(2)_w$ component is referred to as weak isospin, and is also non-abelian. The $\mathrm{U}(1)_Y$ group is referred to as hypercharge, and it is abelian. The SM Lagrangian will contain kinetic terms for each of these three gauge fields:

$$\mathcal{L}_{SM} \supset -\frac{1}{4}(G^A_{\mu\nu})^2 - \frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}(B_{\mu\nu})^2, \tag{2.1}$$

where $G^A_{\mu\nu}$ is the $\mathrm{SU}(3)_c$ field strength tensor (with $A = 1...8$), $W^a_{\mu\nu}$ is the $\mathrm{SU}(2)_w$ field strength tensor (with $a = 1...3$), and $B_{\mu\nu}$ is the $\mathrm{U}(1)_Y$ field strength tensor.

In the SM, the $\mathrm{SU}(3)_c$ symmetry is exact, while the $\mathrm{SU}(2)_w \times \mathrm{U}(1)_Y$ is spontaneously broken by the Higgs mechanism. In order to preserve $\mathrm{SU}(3)_c$, the Higgs field ϕ must transform as a singlet under $\mathrm{SU}(3)_c$; in order to break $\mathrm{SU}(2)_w$ and $\mathrm{U}(1)_Y$, the Higgs field must be charged under these symmetries. The Higgs field is a doublet under $\mathrm{SU}(2)_w$, and has a hypercharge of 1/2. This can be expressed as $\phi = (0, 2, 1/2)$, where the first number corresponds to the $\mathrm{SU}(3)_c$ representation, the second number corresponds to the $\mathrm{SU}(2)_w$ representation, and the third number corresponds to the $\mathrm{U}(1)_Y$ representation. With the inclusion of the Higgs field, the SM Lagrangian will contain:

$$\mathcal{L}_{SM} \supset |D_\mu \phi|^2 + V(\phi), \tag{2.2}$$

where $D_\mu \phi$ is the covariant derivative and V is the potential, given by

$$V(\phi) = \lambda \left(|\phi|^2 - \frac{v^2}{2} \right)^2, \tag{2.3}$$

which is minimized when $|\phi|^2 = v^2/2$. The Higgs field ϕ is a complex doublet:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} h_1 + ih_2 \\ h_0 + ih_3 \end{pmatrix}, \tag{2.4}$$

so we have $\phi^\dagger \phi = 1/2\ (h_0^2 + h_1^2 + h_2^2 + h_3^2)$. We know that $V(\phi)$ is minimized when $\phi^\dagger \phi = v^2/2$, but there are infinitely many ways to satisfy this. We chose $\langle h_0 \rangle = v$, with $\langle h_1 \rangle = \langle h_2 \rangle = \langle h_3 \rangle = 0$, breaking the symmetry. With this choice, the vacuum expectation value for ϕ is:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.5}$$

Plugging equation 2.5 into the Higgs kinetic term $|D_\mu \phi|^2$ gives rise to the mass terms for the massive gauge bosons. Expanding Eq. 2.5 around the minimum, we have

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \tag{2.6}$$

where h is the physical Higgs boson particle.

Next, let us consider the matter fields in the SM. The matter fields and their $\mathrm{SU}(3)_c \times \mathrm{SU}(2) \times \mathrm{U}(1)$ representations are listed in Table 2.1.

TABLE 2.1

MATTER FIELDS IN THE SM.

	$\mathrm{SU}(3)_c$	$\mathrm{SU}(2)_w$	$\mathrm{U}(1)_Y$
Q	3	2	1/6
u	3	0	2/3
d	3	0	-1/3
L	0	2	-1/2
e	0	0	-1

We can write kinetic terms (of the form $i\bar{\psi}\not{D}\psi$) for each of the matter fields. However, because the left and right hand fields transform differently under $\mathrm{SU}(2) \times \mathrm{U}(1)$ (i.e. the left and right hand fields have different $\mathrm{SU}(2) \times \mathrm{U}(1)$ representations), we cannot write Dirac mass terms for the matter fields, since these terms would not be invariant under the SM symmetries. We can, however, use the Higgs field to write Yukawa terms for the matter fields, which gives rise to mass terms and to terms that describe the interactions between the fermions and the Higgs boson.

Putting together Eq. 2.1 (the kinetic terms for the gauge fields), Eq. 2.2 (the kinetic and potential terms for the Higgs field), the kinetic terms for the fermion fields, and the Yukawa terms for the fermion fields, we obtain the full SM Lagrangian.

2.2 Effective field theory

As introduced in Chapter 1, EFT provides a general framework for describing the off-shell effects of heavy new physics as an expansion of higher-dimensional[1] operators. The operators are constructed of products of SM fields and their derivatives. At each order in the expansion, the operators are scaled by powers of Λ, the mass scale of the new physics. All operators of odd dimension violate baryon and/or lepton number [6], so are not considered in this analysis. Operators of dimension six thus represent the leading new physics effects. The dimension-six operators can be expressed in different bases, the most common of which is known as the Warsaw basis [9]. The number of operators at each dimension depends on the flavor symmetries that are assumed. Under the assumption that all three generations may vary independently, there are 2499 operators at dimension six; assuming flavor universality, this number reduces to 59 (assuming lepton number and baryon number conservation) [1].

Adopting a flavor symmetry assumption in between the two extremes, the model presented in Ref. [5] is the EFT model used in this thesis. This model is referred to as the dim6top model; it makes use of the Warsaw basis and provides tree-level modeling for dimension-six operators. Developed to facilitate studies focusing on third-generation effects, the dim6top model has compiled the set of 33 dimension-six operators involving two or more third-generation quarks. In the dim6top model, the operators are assumed to be invariant under $\mathrm{U}(2)_\mathrm{Q} \times \mathrm{U}(2)_\mathrm{u} \times \mathrm{U}(2)_\mathrm{d}$, so the couplings for operators involving third generation quarks may vary independently from the first two generations. While dim6top allows for EFT effects to vary independently for each generation of leptons, this analysis imposes the assumption that the EFT effects impact each lepton generation in the same way.

In this analysis, we aim to include all operators from the dim6top model that

[1]Here dimension refers to the mass dimension of the operator in natural units. The SM operators are of dimension four, so "higher dimensional" refers to operators of dimension greater than four.

TABLE 2.2

LIST OF WILSON COEFFICIENTS INCLUDED IN THE ANALYSIS.

Category	WCs
Four heavy	c_{QQ}^{1}, c_{Qt}^{1}, c_{Qt}^{8}, c_{tt}^{1}
Two light two heavy	c_{Qq}^{31}, c_{Qq}^{38}, c_{Qq}^{11}, c_{Qq}^{18}, c_{tq}^{1}, c_{tq}^{8}
Two heavy two lepton	$c_{Q\ell}^{3(\ell)}$, $c_{Q\ell}^{-(\ell)}$, $c_{Qe}^{(\ell)}$, $c_{t\ell}^{(\ell)}$, $c_{te}^{(\ell)}$, $c_{t}^{S(\ell)}$, $c_{t}^{T(\ell)}$
Two heavy with bosons	$c_{t\varphi}$, $c_{\varphi Q}^{-}$, $c_{\varphi Q}^{3}$, $c_{\varphi t}$, $c_{\varphi tb}$, c_{tW}, c_{tZ}, c_{bW}, c_{tG}

significantly impact processes in which one or more top quarks are produced in association with charged leptons; as listed in Table 2.2, this comes to 26 operators in total. The definitions of the WCs in Table 2.2 and the definitions of the corresponding operators can be found in Table 1 of Ref. [5]. However, for the vertices involving the c_{tG} WC, there is one important difference with respect to the definition in Ref. [5]. In order to allow MadGraph to properly handle the emission of gluons from the vertices involving the c_{tG} WC, an extra factor of the strong coupling is applied to the c_{tG} coefficients (as explained in [8]).

The 26 operators fall into four main categories: operators involving 4 heavy quarks, operators involving two heavy quarks and two light quarks, operators involving two heavy quarks and two leptons, and operators involving two heavy quarks and bosons. The four four-heavy WCs only have significant impacts on the $\mathrm{t\bar{t}t\bar{t}}$ signal process, so these WCs are not included in the modeling of the other five signal processes. The details of the Monte Carlo generation of the signal samples will be discussed in Section 3.2.1.

For each of the six signal processes, we account for diagrams with zero EFT vertices (i.e. the SM contribution) and diagrams with one EFT vertex (i.e. the new physics contribution). The amplitude for each process will thus depend linearly on the WCs, and the cross sections will depend quadratically on the WCs. Since the weight of each generated event corresponds to the event's contribution to the inclusive cross section, each event weight will also depend quadratically on the 26 WCs; the parametrization of the event weights in terms of the WCs is the key concept that allows us to obtain detector-level predictions in terms of the WCs and will be discussed in detail in Section 3.2.2.

CHAPTER 3

DATA AND MONTE CARLO SAMPLES

The data samples used in this analysis are described in Section 3.1. The Monte Carlo (MC) simulated samples are described in Section 3.2, with the generation of the privately produced samples (with the necessary EFT weights) covered in 3.2.1, and the quadratic parametrization of the weights detailed in 3.2.2.

3.1 Data samples and triggers

This analysis uses data from proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the CMS experiment during 2016, 2017, and 2018 with a combination of single, double, and triple lepton triggers. The total integrated luminosity is 137.6 fb^{-1} with an uncertainty of 1.6% fb^{-1} [12]. The set of triggers and luminosity blocks are provided for reference in Appendix A. Since one trigger can be a part of multiple datasets, the overlap between datasets must be accounted for in order to avoid double counting. The following procedure is used:

- An arbitrary order of the datasets from a given year is chosen.
- An event that is from the first dataset (dataset A) is never discarded.
- An event that is from the second dataset (dataset B) is discarded if it passes any of the triggers from dataset A (since it was already accounted for in dataset A).
- An event that is from the third dataset (dataset C) is discarded if it passes any of the triggers from dataset A or dataset B (since it was already accounted for).
- The procedure continues for all of the datasets that are included in the given data-taking period.

3.2 Monte Carlo samples

This analysis aims to study dimension-six EFT effects on processes in which one or more top quarks are produced in association with additional charged leptons; processes which lead to the same multilepton final-state signatures but are not impacted by these EFT operators are backgrounds for this analysis. The expected background contributions are estimated using a combination of simulated samples and data-driven techniques, discussed in Chapter 8 (with the simulated samples used in the background estimation listed in Appendix A). The details of the signal sample generation are described in Section 3.2.1, and 3.2.2 covers the details that are specific to the EFT weights.

3.2.1 Monte Carlo generation of signal samples

The signal processes for this analysis are $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l}\nu$, $\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$, tHq, and $\mathrm{t\bar{t}t\bar{t}}$. The signal samples are produced at leading order (LO) with the MadGraph [3] event generator (version 2.6.5). As discussed in Chapter 2, the dim6top UFO model [5] is used to incorporate the EFT effects. Parton showering and hadronization for the samples is performed with the Pythia generator [19], which also handles the decays of the top quark and the Higgs boson. In order to avoid overlap between the $\mathrm{t\bar{t}l\bar{l}}$ and $\mathrm{t\bar{t}H}$ samples, we specify in the MadGraph process card that the $\mathrm{t\bar{t}l\bar{l}}$ process should not include an intermediate H; the same requirement is made for the $\mathrm{tl\bar{l}q}$ process in order to avoid overlap with tHq. All simulated signal processes are normalized to the latest theoretical cross sections at next-to-leading order (NLO) in QCD, as listed in Table A.5. The private EFT samples produced for this analysis are located at the Notre Dame T3. Tor reference their file paths are listed in Tables A.6, A.7, A.8, and A.9.

For the $\mathrm{t\bar{t}X}$ processes ($\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l}\nu$, and $\mathrm{t\bar{t}l\bar{l}}$), we include an additional final state parton in the matrix element (ME) generation. The inclusion of the additional parton can improve the modeling at high jet multiplicities, and can also significantly impact the dependence of the $t\bar{t}X$ processes on the WCs [8]. The primary factors contributing to the modification of the cross section's EFT dependence are related to the new quark-gluon initiated diagrams that become available when an additional final state parton is included in the ME calculation. Other factors, such as the chiral and color structure of the operator, can also play an important role. The single top processes ($tl\bar{l}q$ and tHq) and the $t\bar{t}t\bar{t}$ sample are not produced with an additional parton. The single top processes have technical complications associated with correctly performing the jet matching between the ME and the parton shower (PS) for t-channel single top processes that currently do not allow a valid matched sample to be produced. In the case of $t\bar{t}t\bar{t}$, an additional parton is not included because the generation of the MadGraph gridpack is very computationally expensive. It would not be feasible to produce enough $t\bar{t}t\bar{t}$ samples to perform a thorough validation of the starting point and matching parameters validation.

Since we are unable to include an additional parton for the single top samples, and in these cases the extra parton may potentially have a significant effect on the high jet multiplicity categories (since these single top processes would not generally produce as many jets as our other signal processes), we apply an additional uncertainty to these processes, described in Chapter 9. This uncertainty is determined by comparing the jet multiplicity distribution of our private EFT samples (reweighted to the SM) against centrally produced NLO samples, listed in Table A.10.

For the samples produced with an additional parton, a matching procedure must be applied to account for the overlap in phase space between the contributions of the ME and parton shower (PS). For this analysis, the matching is implemented using the MLM scheme [4], an event-rejection based approach that matches ME partons to jets clustered by Pythia, discarding events in which the jets are not successfully matched to partons in order to avoid double counting.

It should be noted that the matching procedure can lead to complications when applied to EFT samples; since EFT effects are included in the ME contribution, but not in the PS contribution, it is possible that an inconsistency could arise. Specifically, if an EFT vertex produces a significant soft and collinear contribution, the events removed by the matching procedure will never be replaced by corresponding events generated by the PS, causing this contribution to be missed. However, of the WCs considered in this study, the operator associated with the c_{tG} WC is the most prone to these effects, and its contributions to the soft and collinear regime are suppressed; thus, the phase space overlap with the SM contribution from the PS is small, and the effects of this potential issue are negligible [8].

In addition to the theoretical justification outlined above, we can validate the matching procedure empirically by examining differential jet rate (DJR) distributions for the simulated samples. Additional information about the validation of the DJR distributions may be found in Appendix B, and a more detailed discussion of the validation of matched $\mathrm{t\bar{t}X}$ samples is presented in [8].

As an additional form of validation, our privately produced signal samples (reweighted to the SM) are compared against SM samples that are centrally produced by the CMS collaboration. The details of this comparison are presented in Appendix C.

3.2.2 Parameterization of the predicted yields in terms of the WCs

This section will describe the method through which the predicted yields are parameterized in terms of the WCs. In order to write the predicted yields as a function of the WCs, it is first necessary to understand how the cross section depends on the WCs. Starting with the ME, we can write the amplitude for a given process as the sum of the SM and new physics components:

$$\mathcal{M} = \mathcal{M}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{M}_i, \tag{3.1}$$

where $\mathcal{M}_{\mathrm{SM}}$ is the SM ME, $\mathcal{M}_i$ are the MEs corresponding to the new physics components, and c_i are the WCs. Since the cross section (inclusive or differential) is proportional to the square of the ME, it will depend quadratically on the WCs:

$$\sigma \propto |\mathcal{M}|^2 \propto s_0 + \sum_j^N s_{1j}\frac{c_j}{\Lambda^2} + \sum_j^N s_{2j}\frac{c_j^2}{\Lambda^4} + \sum_{j\neq k}^N s_{3jk}\frac{c_j}{\Lambda^2}\frac{c_k}{\Lambda^2}, \tag{3.2}$$

where s_{jk} are structure constants of the N-dimensonal quadratic function for N WCs. The number of structure constants (K) required to describe an N-dimensional quadratic can be written as the following:

$$K = \frac{(N+1)\cdot(N+1) - (N+1)}{2} + (N+1). \tag{3.3}$$

This analysis considers 26 WCs, so by Eq. 3.3, there are 378 structure constants required to fully describe the 26-dimensional quadratic. In principle, we could solve for these structure constants if the cross section at 378 points in the 26-dimensional WC space were known. However, this would require generating 378 unique simulated samples at 378 unique points in the 26-dimensional WC space. In practice, it would not feasible to generate this many simulated samples.

Instead of attempting to determine the parametrization for the inclusive cross section, we parametrize each event's weight in terms of the WCs. Since each weight corresponds to the event's contribution to the inclusive cross section, the event weight essentially represents a differential cross section, which can be described by a 26-dimensional quadric in terms of the WCs, as written in equation 3.2. In order to determine the 378 structure constants of the event weight's quadratic parameterization, we need to know the event weight at 378 distinct points in the 26 dimensional space. This is feasible to do using the MadGraph event reweighting [14] procedure.

Given an event generated under a specific theoretical scenario, the MadGraph event reweighing procedure computes additional weights associated with the same event under alternative theoretical scenarios. In the case of EFT reweighting, the original theoretical scenario corresponds to a particular point in the 26-dimensional WC space, provided to MadGraph by the user. We refer to this as the "starting point" for the sample. The alternative theoretical scenarios correspond to other distinct points in the 26-dimensional WC space (i.e. other sets of values for the 26 WCs), also provided to MadGraph by the user. From the matrix-element computations, MadGraph calculates the weight at the starting point and at each of the additional reweight points. With at least 378 weights corresponding to 378 independent points in the 26-dimensional WC space, we can solve for the 26 structure constants, and fully determine the 26-dimensional quadric function that describes the event's weight in terms of the WCs.

Once we have obtained each event's 26-dimensional quadratic parametrization $w_i(\vec{c}/\Lambda^2)$, we can find the dependence of any observable bin on the WCs by summing the quadratic parameterizations for each of the events that passes the selection criteria for the given bin. The yield Y for a given bin can thus be written as

$$\begin{aligned} \mathrm{Y}\left(\frac{\vec{c}}{\Lambda^2}\right) &= \sum_i w_i\left(\frac{\vec{c}}{\Lambda^2}\right) \\ &= \sum_i \left(s_{0i} + \sum_j s_{1ij}\frac{c_j}{\Lambda^2} + \sum_j s_{2ij}\frac{c_j^2}{\Lambda^4} + \sum_{j\neq k} s_{3ijk}\frac{c_j}{\Lambda^2},\frac{c_k}{\Lambda^2}\right), \end{aligned} \tag{3.4}$$

where the sum over i corresponds to the sum over all of the events that pass the selection criteria for the given bin. Since the sum of multiple quadratic functions is also quadratic, the yield in each bin will be quadratic in terms of the WCs.

Since we are thus able to write the predicted yield of any observable bin as a function of the 26 WCs, we can obtain detector-level predictions at any arbitrary point in the 26-dimensional EFT space. This is the key enabling concept of this analysis, as it allows for all EFT effects across all analysis bins to be simultaneously accounted for when performing the likelihood fitting with the statistical framework (which will be described in Chapter 10).

We generate all of our signal processes using this procedure. However, we do not include all WCs for all processes (since some of the WCs do not impact all of the processes), so the number of reweight points included in the MC generation varies by sample. The $\mathrm{t\bar{t}t\bar{t}}$ process incorporates the full set of 26 WCs. By Eq. (3.3), a total of 378 weights are required to fully determine the 26-dimensional quadratic parameterization. However, in order to ensure that a good fit can be found, we over-constrain the fit by including approximately 20% more points than the minimum number required, for a total of 454 reweight points. As discussed in Section 2.2, the other five signal samples have a negligible dependence on the four four-heavy operators to which $\mathrm{t\bar{t}t\bar{t}}$ is sensitive, so these samples incorporate only 22 WCs. This means a minimum of 276 reweight points are required to determine the 22-dimensional quadratic fit, but we again ensure the fit is over-constrained by generating additional reweight points, for a total of 332 reweight points for each event.

The MadGraph reweighting procedure is powerful because it allows different regions of EFT phase space to be probed with a single Monte Carlo (MC) sample; however, there is an important caveat to the procedure that should be highlighted. Since MadGraph produces unweighted samples of events, the events generated by MadGraph mainly correspond to phase space occupied by original event. Thus, the reweighting procedure does not work unless the original point in phase space (i.e. the starting point) and the alternative points in phase space (i.e. the reweight point) have some overlap. EFT operators lead to new diagrams that may populate areas in phase space that are not present in the SM, therefore the SM cannot be used as a valid starting point for the reweighting procedure. Instead, a point that is relatively far from the SM should be chosen. Nevertheless, even for non-SM starting points, there is still no guarantee that the chosen point will allow MadGraph to properly reweight to all areas of relevant phase space. Therefore, it is important to validate reweighted samples to ensure that they are able to be consistently reweighted to as much of the relevant phase space as possible. For example, we check that the samples are able to be consistently reweighted to other points in EFT phase space (by comparing against dedicated samples produced at the given point in phase space), as well as checking the distribution of event weights for samples generated at different starting points.

CHAPTER 4

ACCELERATOR AND DETECTOR

4.1 The large hadron collider

4.2 The CMS detector

4.2.1 The solenoidal magnet

4.2.2 The inner tracker

4.2.3 The electromagnetic calorimeter

4.2.4 The hadronic calorimeter

4.2.5 The muon detectors

4.2.6 The trigger systems

CHAPTER 5

OBJECT RECONSTRUCTION

5.1 Particle flow

5.2 Object selection

5.2.1 Leptons

5.2.2 Jets and b-tagging

CHAPTER 6

EVENT SELECTION

6.1 2lss category

6.2 3l category

6.3 4l category

CHAPTER 7

DATA-TO-MC CORRECTIONS

7.1 Pileup reweighting

7.2 Trigger efficiency

7.3 Lepton efficiency

7.4 Jet energy scale and resolution

7.5 b-tag efficiency and mistag rate

7.6 ECAL and muon prefiring correction

7.7 Muon energy scale and resolution

CHAPTER 8

BACKGROUNDS

8.1 Irreducible backgrounds

8.2 Reducible backgrounds

8.2.1 Nonprompt background

8.2.2 Charge misidentification background

CHAPTER 9

SYSTEMATICS

CHAPTER 10

SIGNAL EXTRACTION

Once the event selection has been performed and the selected events are binned according to the differential kinematical distributions (as described in Chapter 6), the next step is to perform the statistical analysis in order to extract the confidence intervals (CIs) for the WCs. In Section 10.1 we will step through the relevant statistical concepts and tools. In Section 10.2, the details of the statistical framework will be be described. Finally, in Section 10.3, one of the challenges of multi-dimensional EFT fits will be discussed, and the workaround developed to mitigate this challenge will be explained.

10.1 Likelihood fitting

Let us start by defining the likelihood L as the probability to have measured the observed number of data events, given some theory, i.e. $L = P(\text{data}|\text{theory})$. The number of observed events should follow a Poisson distribution, with a mean corresponding to the number of predicted events. Since each bin is statistically independent, the likelihood will be given by the product of the Poisson probabilities for all of the bins in the analysis.

For many analyses, the predicted number of events in a given bin i can be written as $\mu s_i + b_i$, where b_i is the expected number of background events, s_i is the expected number signal events (according to the SM prediction), and μ is a free parameter. The μ parameter is usually referred to as the signal strength; it must be greater than or equal to 0, and it is constant across all bins. However, in this analysis, we cannot write the predicted number of events as $\mu s_i + b_i$, because the prediction does not scale linearly with a universal signal strength μ. Rather, the predicted number of events in each bin depends quadratically on the 26 WCs (and the quadratic parameterization is different in each bin), as detailed in Section 3.2.2; we will write this prediction as $\mu(\theta)_i$, where θ are the values of the WCs. This function $\mu(\theta)_i$ represents the prediction in each bin (which in our case is a 26-dimensional quadratic in terms of the WCs), and should not be confused with the signal strength μ. With this notation, we can write the likelihood as follows:

$$L = \prod_{i=1}^{N} \frac{\mu(\theta)_i^{n_i}}{n_i!} e^{-\mu(\theta)_i}, \tag{10.1}$$

where N is the number of bins, n_i is the number of observed events in bin i, and $\mu(\theta)_i$ is the number of predicted events in bin i (as a function of the WCs θ).

In Eq. 10.1, the θ represents the set of 26 dimension-six WCs studied in this analysis. To understand the likelihood's dependence on the WCs, we could in principle perform a 26-dimensional grid scan. To perform the grid scan, we would choose a reasonable range for each WC (based on the estimated sensitivity to the WC), chose a granularity with which to scan, and then proceed to record the likelihood at each point on the 26-dimensional grid. However, this "brute-force" approach scales exponentially with the number of dimensions, and becomes prohibitively expensive when more than a few dimensions are considered. Let us step through an example for 26 dimensions. Even if we chose a very sparse grid with only 5 scan points in each dimension, we would still need to scan 5^{26} points. Assuming ~ 1 hour per scan point (a typical length of time for the fits in this analysis) and 10k CPU cores (a reasonable amount of resources we could utilize with an opportunistic pool such as Notre Dame's CRC), it would take about 17 billion years to perform the scan. This brute force approach is thus not feasible for our analysis.

Instead of a 26-dimensional scan, we perform a 1-dimensional scan for each WC, profiling the other 25 WCs. Continuing to refer to the scanned parameter as θ, let us refer to the profiled parameters as ν. In the 1-dimensional scan, we step along one axis θ in the 26-dimensional space (i.e. we step through a set of values for one WC). At each of the steps along the θ axis, the profiled parameters ν are set to the values that cause the likelihood to be maximized at this given value of the scanned parameter θ. The profile likelihood L_p is thus written as follows:

$$L_p(\theta) = L(\theta, \hat{\hat{\nu}}(\theta)), \tag{10.2}$$

where the double-hat notation denotes the values of the profiled parameters ν that maximize the likelihood for the given θ. The profiled likelihood is thus a function of θ only; it is not a function of ν, since the profiled parameters ν do not freely vary (they are a function of θ). As these scans are only 1-dimensional, they are much less computationally expensive than the 26-dimensional scan described above. We can thus perform a 1-dimensional scan for each of the WCs in order to determine the 1-dimensional profile likelihood for each dimension. In principle we can extend this concept to scan over any number of the WCs (profiling the remaining WCs); however, in practice, the largest number of parameters we can scan is 2 (since even for a 3 dimensional scan, the space is too large to fully explore with our current computational abilities).

Next, we would like to understand how the profile ratio compares to the maximum likelihood as a function of θ. We will refer to the values of θ and ν that globally maximize the likelihood as $\hat{\theta}$ and $\hat{\nu}$, respectively. We can then write the profile likelihood ratio $\lambda_p(\theta)$ as follows:

$$\lambda_p(\theta) = \frac{L(\theta, \hat{\hat{\nu}}(\theta)}{L(\hat{\theta}, \hat{\nu})}, \tag{10.3}$$

where the numerator is the profile likelihood from Eq. 10.2, and the denominator is the value of the likelihood at its global maximum. From the profile likelihood ratio, we can form the test statistic $-2\ln\lambda_p(\theta)$. Wilks' theorem [21] states that $-2\ln\lambda_p(\theta)$ should approach a χ^2 distribution in the limit where the data sample is large, where the degrees of freedom correspond to the number of free parameters in the $\lambda_p(\theta)$. For example, for a 1-dimensional scan, there is 1 degree of freedom. To find the 1-dimensional confidence intervals for a given WC, we would thus need to perform a scan for the WC, finding the $-2\ln\lambda_p(\theta)$ at each scan point; since the $-2\ln\lambda_p(\theta)$ is assumed to follow a χ^2 distribution with 1 degree of freedom, we can read off the 1 and 2 σ confidence intervals by observing where the $-2\ln\lambda_p(\theta)$ crosses 1 and 4, respectively [22].

This section has discussed how the predicted yield in each bin depends on the parameters of interest (the WCs), but the prediction also depends on the systematic uncertainties (enumerated in Chapter 9). The systematic uncertainties are taken into account via additional free parameters in the fit; these degrees of freedom are referred to as nuisance parameters. When finding the profile likelihood $L(\theta, \hat{\hat{\nu}}(\theta))$, the nuisance parameters are profiled (i.e. they may be included in the ν in Eq. 10.3).

10.2 Statistical framework

The the CMS Higgs `Combine` software tool [17] is used to perform the likelihood fits. The `Combine` tool uses the `ROOT` framework's `RooFit` tools [20] and the `MINUIT2` software library [10] to numerically minimize the negative log of the profile likelihood function described in Section 10.1. As discussed in Section 3.2.2, the expected yield in each bin is parameterized as a 26-dimensional quadratic in terms of the WCs, as given in Eq. 3.4. In order for `Combine` to calculate the profile likelihood described in Section 10.1, the quadratic dependence must be made known to `Combine`. In principle, a template histogram could be defined for each of the 378 structure constants of the 26-dimensional quadratic, with the normalizations of the templates set by the

`PhysicsModel`. However, in practice this is not possible with `Combine`, since the interference terms of the quadratic may be negative, and `Combine` does not handle histograms with negative yields.

To work around this challenge, we use the approach developed in [13]. With this approach, the 378 terms of the quadratic parametrization are rearranged into 378 linear combinations of the original terms, defined such that each term is positive by construction. We can then create a template histogram for each of the rearranged terms, encoding the appropriate normalization of each histogram in the `Combine` `PhysicsModel`. The template histograms and normalizations encoded in the `PhysicsModel` contain the full description of the 26-dimensional quadratic function, so `Combine` is able to appropriately handle this dependence while performing the likelihood fits.

The expected yields also depend on the experimental and theoretical systematic uncertainties (enumerated in Chapter 9), the effects which are taken into account by a set of nuisance parameters. As mentioned in Section 10.1, the nuisance parameters are profiled in the likelihood fit. The systematic uncertainties may affect either the normalization of the template histograms, or both the normalization and the shape of the template histograms. The former are accounted for via rate systematics in `Combine`, and the latter are accounted for via shape systematics in `Combine`. The systematic uncertainties that affect the shape of the template histograms also cary a 26-dimensional quadratic dependence on the WCs, which is accounted for in the same way as the nominal templates.

10.3 Navigating false minima

EFT fits (especially multi-dimensional EFT fits) may lead to features in the likelihood surface; these features may be influenced by interference between the WCs and the SM, by the interference among the WCs, or by deviations or fluctuations in the data. If these features include local minima, the profile fits may be susceptible to incorrectly identifying a local minimum as the true minimum (i.e. the fit may become "stuck" in the local minimum). Cases where the fit becomes "stuck" in local minima may lead to false best fit points, discontinuities in the negative log likelihood (NLL) scans, and inaccurate confidence intervals.

Symptoms of this issue had been observed in the predecessor to this analysis (Ref. [18]) in the form of discontinuities in the NLL values obtained in the 1-dimensional scans. To work around this issue, Ref. [18] performed 2-dimensional scans for pairs of WCs that had been identified as problematic, avoiding the discontinuities by making use of the NLL values obtained in the 2-dimensional scans. However, this approach is not only computationally expensive, but it also does not guarantee that the correct minima will be found (as the fit may encounter similar challenges with false minima while profiling the remaining $n-2$ WCs); a more general approach would be beneficial. Since local minima can arise as a result of the interference terms in the n-dimensional quadratic parameterizations, the challenge of navigating local minima seems to be an inherent feature of multi-dimensional EFT likelihood fitting. As the EFT community continues to explore simultaneous fits to larger sets of WCs, this pernicious issue may become increasingly problematic. For these reasons, one of the intermediate goals of this analysis was to develop a more general approach to the navigation of local minima within the `Combine` framework.

Before developing an approach to address these issues, we wanted to first gain a better understanding of the underlying cause. To this end, we worked to reproduce the issue in a much simpler case. In this simplified model, we only considered two WCs ($c^{-}_{\varphi Q}$ and c_{tG}). We then performed a profiled likelihood fit, scanning over $c^{-}_{\varphi Q}$ and profiling c_{tG}; in other words, we asked the fit to step through a set of given points for $c^{-}_{\varphi Q}$, and at each of those points to find the value of c_{tG} that would minimize the NLL. The result of this likelihood fit is shown in the lefthand side of Figure 10.1.

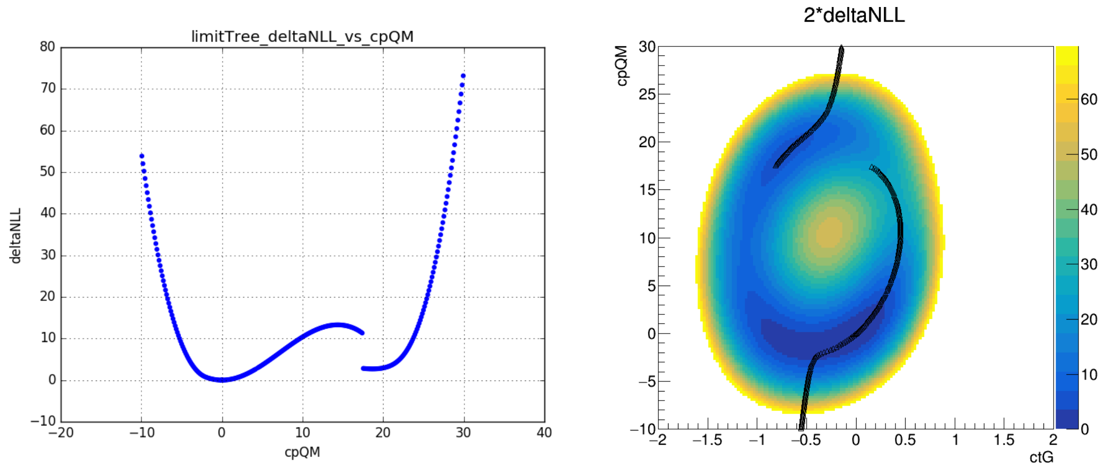


Figure 10.1. Left: NLL for the 1-dimensional profiled fit. The $c^{-}_{\varphi Q}$ parameter is scanned while c_{tG} is profiled. Right: NLL for the 2-dimensional scan. Here both $c^{-}_{\varphi Q}$ and c_{tG} are scanned. The color scale shows the NLL at each of the 2-dimensional scan points. The black overlaid points show the path of the 1-dimensional profiled fit.

At a value of approximately $c^{-}_{\varphi Q} = 17$, a discontinuity is observed. To understand this discontinuity, we performed a 2-dimensional scan over both WCs. This allows us to see the complete picture of the space, helping us to visualize why the profiled fit fails. The 2-dimensional scan is shown on the righthand side of the Figure 10.1, with the path of the 1-dimensional profiled fit overlaid in black points. In the 1-dimensional profiled fit, $c^{-}_{\varphi Q}$ was scanned while c_{tG} was profiled, meaning that for every point along the y direction (i.e. the $c^{-}_{\varphi Q}$ direction) of the 2-dimensional scan, the fit profiles along the x direction (i.e. the c_{tG} direction) in order to find the c_{tG} value at which the NLL (represented by the color scale) is minimized. Following the path of the profiled fit (the overlaid black points), we see that the fit was correctly identifying the c_{tG} point that minimized the NLL from $c^{-}_{\varphi Q} = 0$ until approximately $c^{-}_{\varphi Q} = 5$. At this point, the true minimum lies on the left of the“hill” in the NLL, but the fit continues around the right of this “hill”, subsequently incorrectly identifying these local minima as the best fit points. Once the scan reaches approximately

$c_{\varphi Q}^{-} = 17$, the fit suddenly jumps to the deeper minimum on the left side of the "hill", resulting in the discontinuity observed in the 1-dimensional NLL plot.

To avoid discontinuities, the ideal solution would be to perform a simultaneous scan over all parameters. However, as discussed in Section 10.1, this approach scales exponentially with the number of parameters, and is infeasible for the case of 26 parameters of interst. As an alternative approach, we introduced an element of randomness into the fit in order to sample from the 26-dimensional space. To motivate this approach, let us revisit the profiled fit discussed above. During a fit, `Combine` always uses the same starting value for the profiled parameters; this is useful for reproducibility, but it means that if the starting point happens to be near a local minimum (where the global minimum lies on the other side of a "hill" in NLL), the fit will always find the local minimum, and never find the correct global minim. If we instead allow the starting point for the profiled fit to be chosen randomly, the starting point will sometimes lie on the other side of the "hill" in NLL, allowing the fit to find the correct minimum.

To test the random starting point method, we modified the `Combine` tool to incorporate random starting points for the profiled parameters. After first generating a list of random starting points, the modified version of the `Combine` script loops through the random starting points, finding the NLL at each, and keeping track of the point that gives rise to the lowest NLL. After trying each of the random starting points, the point that produced the lowest NLL is taken to be the profiled value of the parameter at that point. Applying this approach to the 2-dimensional case described above, the NLL is found to be continuous, as shown in Figure 10.2, where the profiled fit finds the correct global minimum at each scan point; for example, at approximately $c_{\varphi Q}^{-}$ =5, the fit is able to identify the deeper minimum on the left side of the "hill" in NLL, jumping to that minimum immediately instead of erroneously continuing around the right side of the "hill" (as had been observed in Figure 10.1).

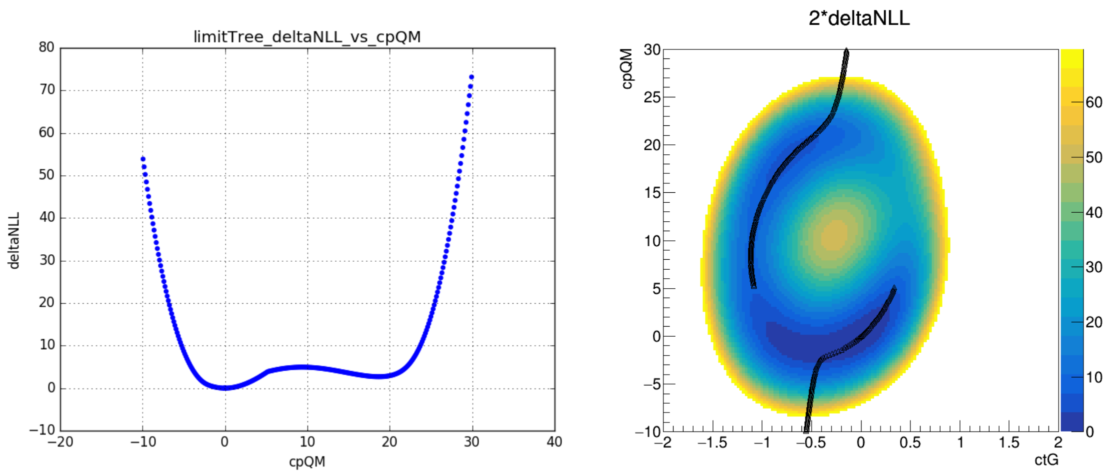


Figure 10.2. Left: NLL for the 1-dimensional profiled fit after the implementation of the random starting point approach, where $c^{-}_{\varphi Q}$ is scanned and c_{tG} is profiled. Right: NLL for the 2-dimensional scan over $c^{-}_{\varphi Q}$ and c_{tG}. The color scale shows the NLL at each of the 2-dimensional scan points. The black overlaid points show the path of the 1-dimensional profiled fit after the implementation of the random starting point approach.

After the random starting point approach was shown to be successful in the simple 2-dimensional case, we generalized the method and tested it on the 16-dimensional fits from [18]. Although the number of random starting points required in order to obtain a smooth NLL curve was larger in the 16-dimensional case than in the 2-dimensional case (greater than 50 as opposed to less than 10), the method successfully avoided discontinuities in NLL. The results of this test are shown in Appendix D.

Although this approach is computationally feasible, it is still relatively computationally expensive (as the likelihood fit must be run m times for each scan point, where m is the number of random starting points). It would thus be interesting to optimize the approach by considering methods of sampling the space more efficiently. For example, one idea would be to first identify a set of distinct local minima in the space, and using a set of points from these local minima as the starting values instead of choosing the starting points randomly from the full space.

CHAPTER 11

RESULTS

CHAPTER 12

SUMMARY

APPENDIX A

REFERENCE INFORMATION FOR DATA AND MC SAMPLES

This appendix includes additional details about the data and MC samples used in this analysis.

Table A.1 lists the CMS JSON files that specify the luminosity blocked used in the analysis. The lumi blocks listed in the files (known as the "golden JSONs") exclude data that is effected by known detector issues. The triggers used for the 2016 data are listed in Table A.2, the triggers used for the 2017 data are listed in Table A.3, and the triggers used for the 2018 data are listed in Table A.4.

The privately generated signal samples for the UL16, UL16APV, UL17, and UL18 periods are listed in Tables A.6, A.7, A.8, and A.9. Table A.5 lists the NLO cross sections used to normalize the signal samples. The files are located at the Notre Dame T3.

Table A.10 lists the central samples used for comparison against our privately produced $\mathrm{tl\bar{l}q}$ EFT samples in order to calculate the additional systematic uncertainty that is applied to the single top samples (for which we are unable to include an additional parton in the matrix element), as described in Chapter 9.

The centrally produced background samples (CMSSW_10_6_26) used in this analysis are listed in Tables A.11, A.12, A.13, and A.14. The first section of the tables lists the samples for the processes for which we use the simulation to estimate the contribution. The second section the samples for processes that are relevant to control region, or for contributions that are estimated from data. The "TTJets*" sample was only used for the estimation of the charge-flip contributions.

TABLE A.1

JSON FILES WITH CERTIFIED LUMINOSITY BLOCKS USED FOR EACH DATA-TAKING YEAR.

Year	Golden JSON	Int. Lumi (fb^{-1})
2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON	36.33
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON	41.48
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON	59.83

TABLE A.2

TRIGGERS USED TO RECORD THE 2016 DATA.

Dataset	2016 Triggers
SingleMuon	`IsoMu24`
	`IsoTkMu24`
	`IsoMu22_eta2p1`
	`IsoTkMu22_eta2p1`
	`IsoMu22`
	`IsoTkMu22`
	`IsoMu27`
SingleElectron	`Ele27_WPTight_Gsf`
	`Ele25_eta2p1_WPTight_Gsf`
	`Ele27_eta2p1_WPLoose_Gsf`
DoubleMuon	`Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ`
	`Mu17_TrkIsoVVL_Mu8_TrkIsoVVL`
	`Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL`
	`Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`
	`TripleMu_12_10_5`
DoubleEG	`Ele23_Ele12_CaloIdL_TrackIdL_IsoVL`
	`Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ`
	`Ele16_Ele12_Ele8_CaloIdL_TrackIdL`
MuonEG	`Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL`
	`Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ`
	`Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ`
	`Mu8_DiEle12_CaloIdL_TrackIdL`
	`Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL`
	`Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ`
	`Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL`
	`Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ`
	`DiMu9_Ele9_CaloIdL_TrackIdL`

TABLE A.3

TRIGGERS USED TO RECORD THE 2017 DATA.

Dataset	2017 Triggers
SingleMuon	IsoMu24
	IsoMu27
SingleElectron	Ele32_WPTight_Gsf
	Ele35_WPTight_Gsf
DoubleMuon	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
	TripleMu_12_10_5
DoubleEG	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	Ele16_Ele12_Ele8_CaloIdL_TrackIdL
MuonEG	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu8_DiEle12_CaloIdL_TrackIdL
	Mu8_DiEle12_CaloIdL_TrackIdL_DZ
	DiMu9_Ele9_CaloIdL_TrackIdL_DZ

TABLE A.4

TRIGGERS USED TO RECORD THE 2018 DATA.

Dataset	2018 Triggers
SingleMuon	IsoMu24
	IsoMu27
EGamma	Ele32_WPTight_Gsf
	Ele35_WPTight_Gsf
	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	Ele16_Ele12_Ele8_CaloIdL_TrackIdL
DoubleMuon	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
	TripleMu_12_10_5
MuonEG	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu8_DiEle12_CaloIdL_TrackIdL
	Mu8_DiEle12_CaloIdL_TrackIdL_DZ
	DiMu9_Ele9_CaloIdL_TrackIdL_DZ

TABLE A.5

THEORETICAL CROSS SECTIONS USED FOR NORMALIZING THE SIGNAL SIMULATION SAMPLES.

Process	cross section (pb)	order
$t\bar{t}H$	0.2151	NLO
$t\bar{t}l\bar{l}$	0.281	NLO
$t\bar{t}l\nu$	0.2352998	NLO
$tl\bar{l}q$	0.0758	NLO
tHq	0.07096	NLO
$t\bar{t}t\bar{t}$	0.012	NLO

TABLE A.6

PRIVATELY PRODUCED UL16 SIGNAL SAMPLES.

Process	Xsec (pb)	Events	Location
$t\bar{t}H$	0.2151	8.0M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\nu$	0.2353	9.1M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\bar{l}$	0.281	8.1M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
$tl\bar{l}q$	0.0758	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
tHq	0.07096	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}t\bar{t}$	0.0120	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16/Round1/Batch1/naodOnly_step/v2/nAOD_step_tttt_FourtopsMay3v1_run0

TABLE A.7

PRIVATELY PRODUCED UL16APV SIGNAL SAMPLES.

Process	Xsec (pb)	Events	Location
$t\bar{t}H$	0.2151	8.0M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\nu$	0.2353	9.1M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\bar{l}$	0.281	8.1M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
$tl\bar{l}q$	0.0758	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
tHq	0.07096	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}t\bar{t}$	0.0120	7.5M	/store/user/kmohrman/FullProduction/FullR2/UL16APV/Round1/Batch1/naodOnly_step/v2/nAOD_step_tttt_FourtopsMay3v1_run0

TABLE A.8

PRIVATELY PRODUCED UL17 SIGNAL SAMPLES.

Process	Xsec (pb)	Events	Location
$t\bar{t}H$	0.2151	15.8M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch1/naodOnly_step/v4/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch2/naodOnly_step/v3/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch3/naodOnly_step/v4/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\nu$	0.2353	18.0M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch1/naodOnly_step/v4/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch2/naodOnly_step/v3/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch4/naodOnly_step/v4/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\bar{l}$	0.281	16.0M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch1/naodOnly_step/v4/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch2/naodOnly_step/v3/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch3/naodOnly_step/v4/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
$tl\bar{l}q$	0.0758	14.7M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch1/naodOnly_step/v4/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch2/naodOnly_step/v3/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch3/naodOnly_step/v4/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
tHq	0.07096	14.8M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch1/naodOnly_step/v4/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch2/naodOnly_step/v3/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch3/naodOnly_step/v4/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}t\bar{t}$	0.0120	15.0M	/store/user/kmohrman/FullProduction/FullR2/UL17/Round1/Batch4/naodOnly_step/v2/nAOD_step_tttt_FourtopsMay3v1_run0

TABLE A.9

PRIVATELY PRODUCED UL18 SIGNAL SAMPLES.

Process	Xsec (pb)	Events	Location
$t\bar{t}H$	0.2151	15.6M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch1/naodOnly_step/v5/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch2/naodOnly_step/v2/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch3/naodOnly_step/v2/nAOD_step_ttHJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\nu$	0.2353	18.1M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch1/naodOnly_step/v5/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch2/naodOnly_step/v2/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch4/naodOnly_step/v2/nAOD_step_ttlnuJet_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}l\bar{l}$	0.281	16.2M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch1/naodOnly_step/v5/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch2/naodOnly_step/v2/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch3/naodOnly_step/v2/nAOD_step_ttllNuNuJetNoHiggs_all22WCsStartPtCheckdim6TopMay20GST_run0
$tl\bar{l}q$	0.0758	15.0M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch1/naodOnly_step/v5/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch2/naodOnly_step/v2/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch3/naodOnly_step/v2/nAOD_step_tllq4fNoSchanWNoHiggs0p_all22WCsStartPtCheckV2dim6TopMay20GST_run0
tHq	0.07096	15.0M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch1/naodOnly_step/v5/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch2/naodOnly_step/v2/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
			/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch3/naodOnly_step/v2/nAOD_step_tHq4f_all22WCsStartPtCheckdim6TopMay20GST_run0
$t\bar{t}t\bar{t}$	0.0120	14.9M	/store/user/kmohrman/FullProduction/FullR2/UL18/Round1/Batch4/naodOnly_step/v2/nAOD_step_tttt_FourtopsMay3v1_run0

TABLE A.10

CENTRAL TZQ SAMPLES USED FOR CALCULATING THE ADDITIONAL SYSTEMATIC UNCERTAINTY THAT IS APPLIED TO THE SINGLE TOP SAMPLES.

Year	Sample
UL16APV	/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv6-PUMoriond17_Nano25Oct2019_102X_mcRun2_asymptotic_v7-v1/NANOAODSIM
UL16	/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
UL17	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
UL18	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM

TABLE A.11

LIST OF UL16APV BACKGROUND SAMPLES.

UL16APV Background Samples	Xsec (pb)
/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	3.697
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	87.31
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	364.351
/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.082
/TWZToLL_thad_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Whad_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	0.0015
/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	12.178
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-[v1,ext1-v1]/NANOAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-[v1,ext1-v1]/NANOAODSIM	0.1651
/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	5.2843
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-[v1,ext1-v1]/NANOAODSIM	0.05565
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	1.256
/GluGluToContinToZZTo2e2mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2nu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v3/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo4e_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.00159
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-[v1,ext1-v1]/NANOAODSIM	0.01398
/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	55.78
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	18610.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	6025.2
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	3.68
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	136.02
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	80.95
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	35.85
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM	831.76
/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	61526.7

TABLE A.12

LIST OF UL16 BACKGROUND SAMPLES.

UL16 Background Samples	Xsec (pb)
/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	3.697
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	87.31
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	364.351
/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.082
/TWZToLL_thad_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Whad_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.0015
/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	12.178
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-[v1,ext1-v1]/NANOAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-[v1,ext1-v1]/NANOAODSIM	0.1651
/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	5.2843
/WZTo3LNu_mllmin4p0_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2	5.2843
/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM	0.2353
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-[v1,ext1-v1]/NANOAODSIM	0.05565
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	1.256
/GluGluToContinToZZTo2e2mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2nu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.00319
/GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.00319
/GluGluToContinToZZTo4e_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	0.00159
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-[v1,ext1-v1]/NANOAODSIM	0.01398
/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	55.78
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	18610.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	6025.2
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	3.68
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	136.02
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	80.95
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	35.85
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	831.76
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	61526.7

TABLE A.13

LIST OF UL17 BACKGROUND SAMPLES.

UL17 Background Samples	Xsec (pb)
/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	3.697
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	87.31
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	364.351
/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.082
/TWZToLL_thad_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Whad_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.0015
/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	12.178
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-[v1,ext1-v2]/NANOAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.1651
/WZTo3LNu_mllmin4p0_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	5.2843
/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	0.2353
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-[v1,ext1-v2]/NANOAODSIM	0.05565
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	1.256
/GluGluToContinToZZTo2e2mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2nu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo4e_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	0.00159
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	0.01398
/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	55.78
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	18610.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	6025.2
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	3.68
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	136.02
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	80.95
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM	35.85
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	831.76
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM	61526.7

TABLE A.14

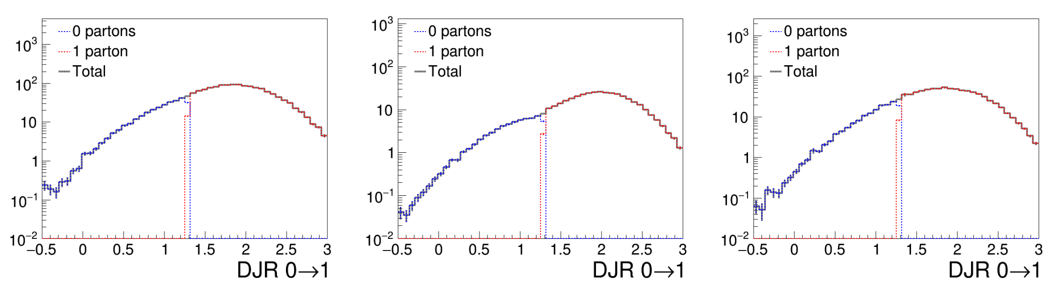
LIST OF UL18 BACKGROUND SAMPLES.

UL18 Background Samples	Xsec (pb)
/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-[v1,ext1-v1]/NANOAODSIM	3.697
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	87.31
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	364.351
/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.082
/TWZToLL_thad_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Whad_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.003004
/TWZToLL_tlept_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.0015
/WWTo2L2NuWWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	12.178
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.1651
/WZTo3LNu_mllmin4p0_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	5.2843
/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.2353
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-[v1,ext1-v2]/NANOAODSIM	0.05565
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	1.256
/GluGluToContinToZZTo2e2mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2nu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2e2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00319
/GluGluToContinToZZTo4e_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00159
/GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	0.00159
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	0.01398
/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	55.78
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	18610.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	6025.2
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	3.68
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	136.02
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	80.95
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM	35.85
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	831.76
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM	61526.7

APPENDIX B

VALIDATION OF LEADING ORDER MATCHING PROCEDURE

The differential jet rate (DJR) can be used as a method of validation for the LO matching procedure [2, 11]. For the k_T algorithm, the DJR histogram represent the distribution of k_T values for which an n jet event transitions to an $n+1$ jet event. A discontinuity in the transition between the n and $n+1$ curves would indicate that there is a mismatch in the overlapping regions of phase space, while a smooth transition is an indication that the ME generator and PS generator are working together to properly fill the phase space without any gaps or double counting. For the matched samples generated for this analysis, we observe smooth DJR plots, as shown in Figure B.1.



(a) (b) (c)

Figure B.1. DJR histograms for matched samples $t\bar{t}H$ (a), $t\bar{t}l\bar{l}$ (b), and $t\bar{t}l\nu$ (c).

In Figure B.1, all WCs have been set to non-zero values. The x axis shows the log base 10 of the scale at which an n jet event transitions into an $n+1$ jet event. The line labeled “0 partons” refers to the contribution from the parton shower, while the line labeled “1 parton” refers to the contribution from the matrix element. The line labeled “Total” is the sum of the two contributions. The smooth transition indicates that the matching scales have allowed the matrix element generator and parton shower to smoothly fill the overlapping phase space.

For the $\mathrm{t\bar{t}X}$ samples generated for this analysis, the matching scales used with MadGraph and Pythia (i.e. the `xqcut` and `qQut`) were 10 and 20, respectively. A more detailed description of the meaning of these matching scales is provided in [8].

APPENDIX C

COMPARISON OF PRIVATELY GENERATED MC SAMPLES TO CENTRALLY GENERATED MC SAMPLES

This analysis uses privately generated LO MC samples for all signal processes, following the same approach as Ref. [18], which is referred to as TOP-19-001. This analysis uses the MG reweighting procedure to incorporate the effects of our 26 WCs into the samples, as explained in Chapter 3. The samples can be reweighted to any arbitrary point in the EFT space, including the SM point (i.e. the point where all WCs are equal to 0). When a sample is reweighted to the SM, it should equivalent (within uncertainties) to a sample that was generated at the SM. In order to verify that the samples produced for this analysis provide a good description at the SM, we can compare the predictions to centrally produced SM samples. Note that some differences are expected in these comparisons, since the central samples are NLO and the privately produced samples are LO. The central samples used for this comparison are listed in Table C.1.

The validation for each signal sample is discussed in more detail in the subsections of this appendix, and briefly summarized in the following bullets:

- $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l}\nu$, $\mathrm{t\bar{t}t\bar{t}}$: These processes are discussed in sections C.1, C.2, and C.3 (respectively). For these processes, there is generally good agreement between privately produced LO samples and centrally produced NLO samples.
- $\mathrm{t\bar{t}l\bar{l}}$: There is some tension for this process, as shown in the plots in section C.4. However, we do not believe this represents a problem for this analysis. The MC validation studies for TOP-19-001 (which involved all pre-UL samples) also showed this tension. While the studies for this analysis (which of course involves all UL samples) have found a somewhat larger disagreement, this seems

TABLE C.1

CENTRAL SAMPLES USED FOR COMPARISON AGAINST OUR PRIVATELY PRODUCED SAMPLES.

Year	Sample
2017	/TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
UL16	/ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM
UL16	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM
UL16	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM
UL16	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM
UL16	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM
UL16APV	/ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM
UL16APV	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM
UL16APV	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM
UL16APV	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM
UL16APV	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM
UL17	/ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM
UL17	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM
UL17	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM
UL17	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM
UL17	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM
UL18	/ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM
UL18	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM
UL18	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM
UL18	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM
UL18	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM

to be due to a change in the default shower starting scale for central UL samples. There does not seem to be any reason to believe that this change represents in improvement in the modeling. For these reasons, it is believed that the current modeling of this process is sufficient (as in TOP-19-001).

- $\mathrm{tl\bar{l}q}$: This process (along with tHq) is discussed in section C.5. The comparison for this sample should be handled carefully because of the fact that we cannot include an additional parton in the matrix element (as explained in Section 3.2.1). However, we already apply an additional uncertainty (derived in N_{jets}) to account for this (as discussed in Chapter 9). The uncertainty covers the discrepancy for the differential distributions used in the analysis, so the current modeling is believed to be sufficient.

C.1 Summary of comparisons for the $\mathrm{t\bar{t}H}$ sample

For this process, there is good agreement between privately produced LO samples and centrally produced NLO samples.

C.2 Summary of comparisons for the $\mathrm{t\bar{t}l}\nu$ sample

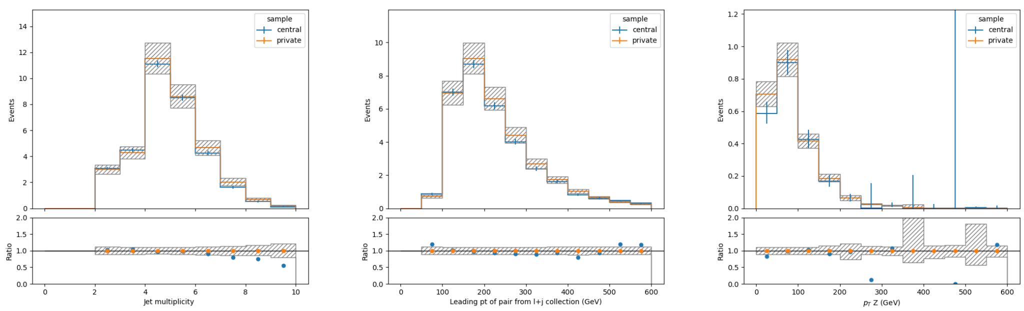
For this process, there is good agreement between privately produced LO samples and centrally produced NLO samples.

C.3 Summary of comparisons for the $\mathrm{t\bar{t}t\bar{t}}$ sample

For this process, there is good agreement between privately produced LO samples and centrally produced NLO samples.

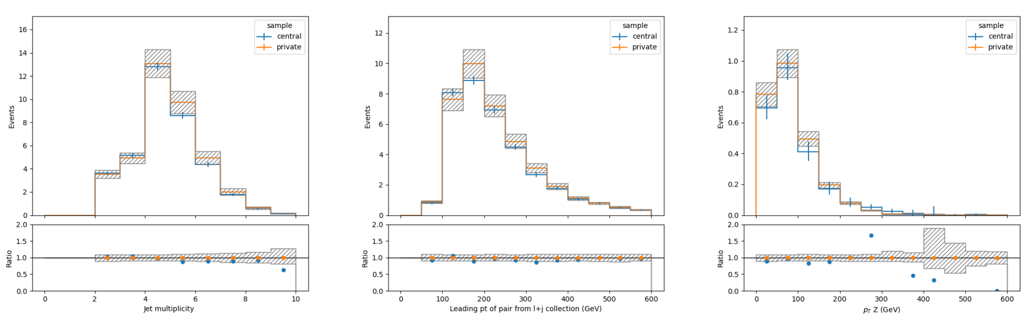
C.4 Summary of comparisons for the $\mathrm{t\bar{t}l\bar{l}}$ sample

There is some tension between the private LO and central NLO samples for this process. However, in TOP-19-001, there was also tension between our private LO samples and the NLO central samples. Unlike the central $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}l}\nu$ samples, the central $\mathrm{t\bar{t}l\bar{l}}$ sample does not explicitly include an extra parton in the matrix element, and we believe it is possible that some portion of the disagreement may be linked



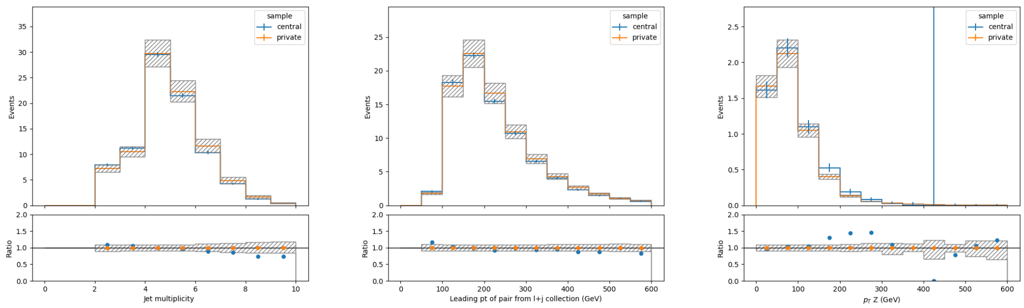
(a) (b) (c)

Figure C.1. RECO level comparison for UL16 $t\bar{t}$H. This plot shows the privately produced LO samples (reweighted to the SM) and the centrally produced NLO samples (datasets used for the central samples are listed in Table C.1). For this comparison, we have summed over all selection categories in the SR. The shaded band represents the systematic uncertainties for the private sample.



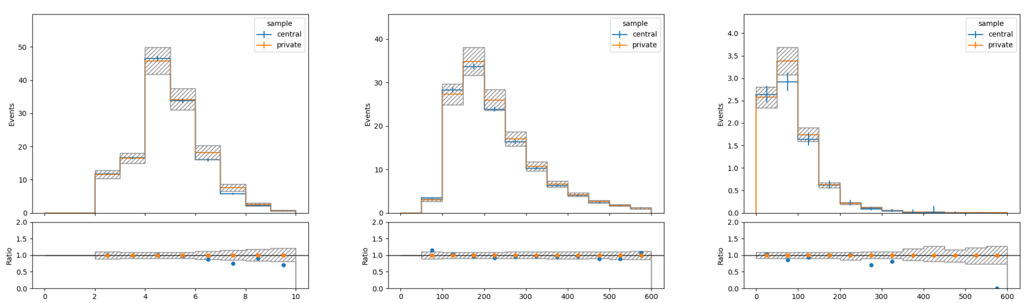
(a) (b) (c)

Figure C.2. RECO level comparison for UL16APV $t\bar{t}$H. All other relevant details are the same as described in Figure C.1.



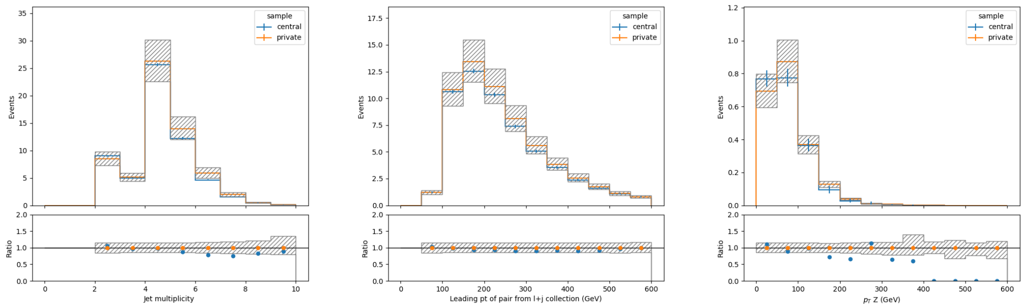
(a) (b) (c)

Figure C.3. RECO level comparison for UL17 $t\bar{t}H$. All other relevant details are the same as described in Figure C.1.



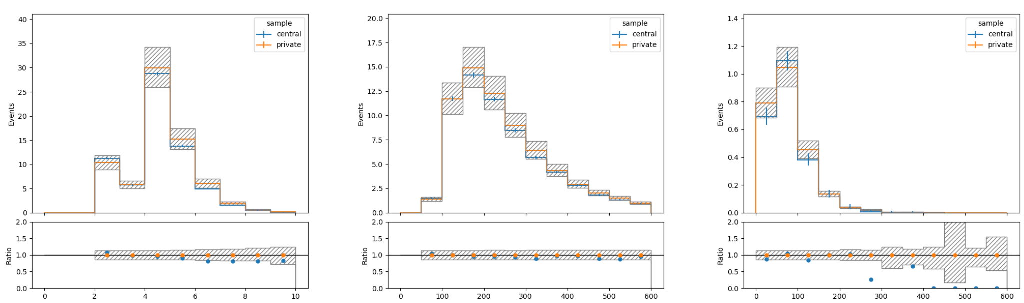
(a) (b) (c)

Figure C.4. RECO level comparison for UL18 $t\bar{t}H$. All other relevant details are the same as described in Figure C.1.



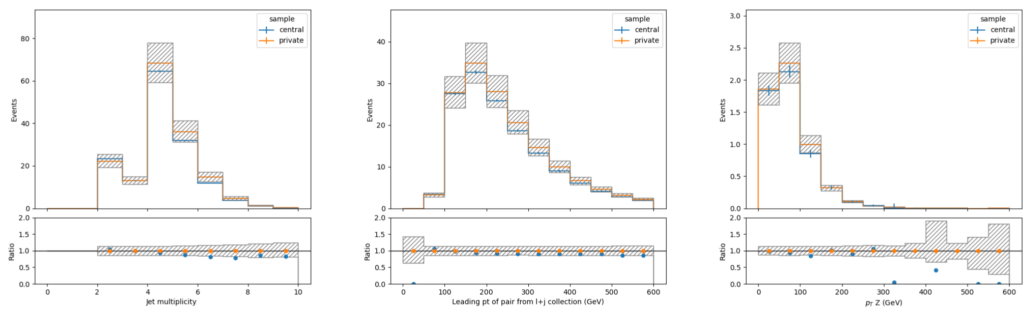
(a) (b) (c)

Figure C.5. RECO level comparison for UL16 $t\bar{t}l\nu$. All other relevant details are the same as described in Figure C.1.



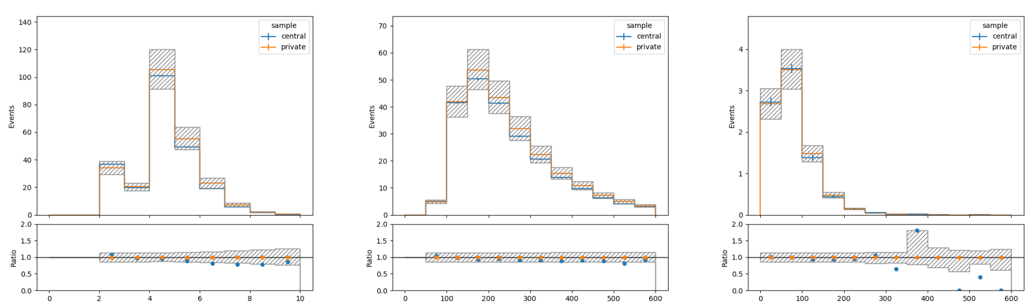
(a) (b) (c)

Figure C.6. RECO level comparison for UL16APV $t\bar{t}l\nu$. All other relevant details are the same as described in Figure C.1.



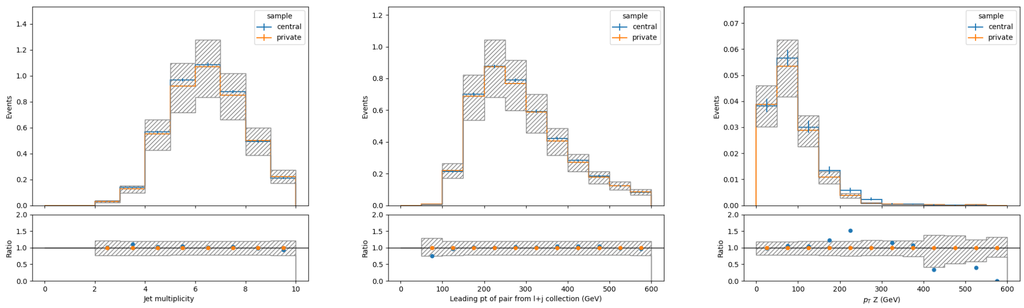
(a) (b) (c)

Figure C.7. RECO level comparison for UL17 $t\bar{t}l\nu$. All other relevant details are the same as described in Figure C.1.



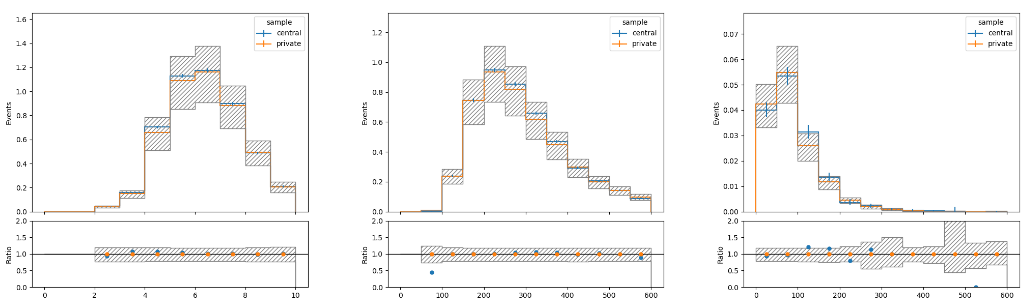
(a) (b) (c)

Figure C.8. RECO level comparison for UL18 $t\bar{t}l\nu$. All other relevant details are the same as described in Figure C.1.



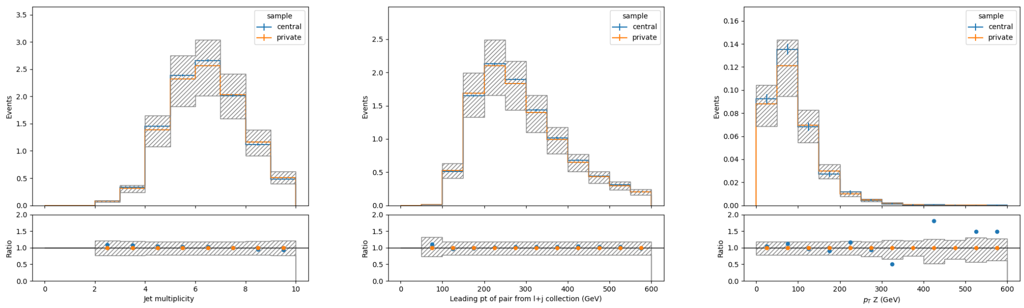
(a) (b) (c)

Figure C.9. RECO level comparison for UL16 $\mathrm{t\bar{t}t\bar{t}}$. All other relevant details are the same as described in Figure C.1.



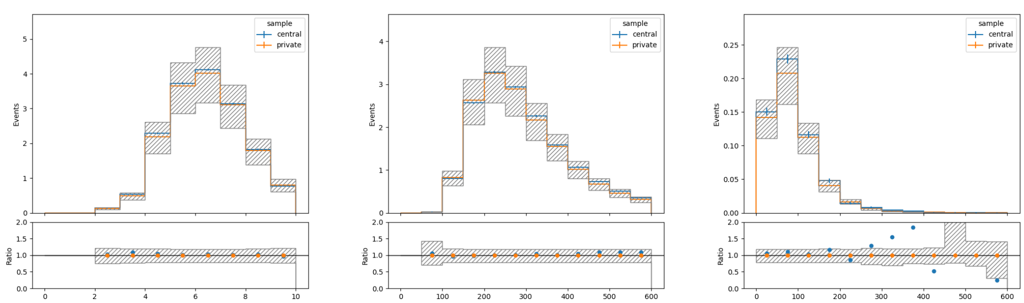
(a) (b) (c)

Figure C.10. RECO level comparison for UL16APV $\mathrm{t\bar{t}t\bar{t}}$. All other relevant details are the same as described in Figure C.1.



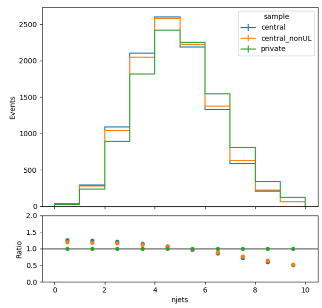
(a) (b) (c)

Figure C.11. RECO level comparison for UL17 $t\bar{t}t\bar{t}$. All other relevant details are the same as described in Figure C.1.

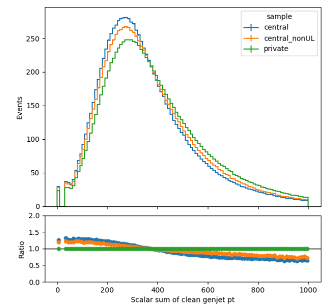


(a) (b) (c)

Figure C.12. RECO level comparison for UL18 $t\bar{t}t\bar{t}$. All other relevant details are the same as described in Figure C.1.



(a)



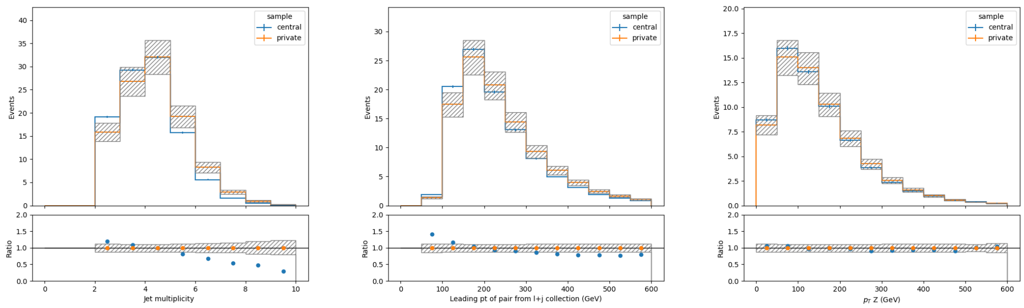
(b)

Figure C.13. GEN level comparison for 2017 and UL17 $\mathrm{t\bar{t}l\bar{l}}$ for the N_{jets} (a) and H_T (b). Some basic jet cleaning has been applied. As can be seen in the plots, the central UL sample has changed in comparison to the central pre-UL samples (and this change happens to make the tension with the private $\mathrm{t\bar{t}l\bar{l}}$ sample somewhat worse). As discussed in the text, the change in the central sample seems to be due to a change in the default shower starting scale, and does not seem to represent an improvement in the modeling of the $\mathrm{t\bar{t}l\bar{l}}$ process.

to this difference. The level of agreement between our private LO samples and the central NLO samples was deemed acceptable for TOP-19-001.

From TOP-19-001 (which used pre-UL samples) to this analysis (which uses all UL samples), our privately produced LO $\mathrm{t\bar{t}l\bar{l}}$ samples have remained consistent. However, there has been a change in the central $\mathrm{t\bar{t}l\bar{l}}$ sample from pre-UL to UL. The change is apparent primarily in jet-related variables, e.g. N_{jets} or H_T (where H_T is defined as the scalar sum of the p_T of all jets in the event). This change seems to be caused by a change in the MadGraph default shower starting scale that was implemented in MadGraph version 2.5.3. Based on discussions with the experts, it seems this change does not necessarily represent an improvement in the modeling of the $\mathrm{t\bar{t}l\bar{l}}$ process. A GEN level comparison of the pre-UL and UL central $\mathrm{t\bar{t}l\bar{l}}$ samples (along with the private UL $\mathrm{t\bar{t}l\bar{l}}$ sample) is shown in Figure C.13.

This change in the central UL samples (which seems to have been caused by the



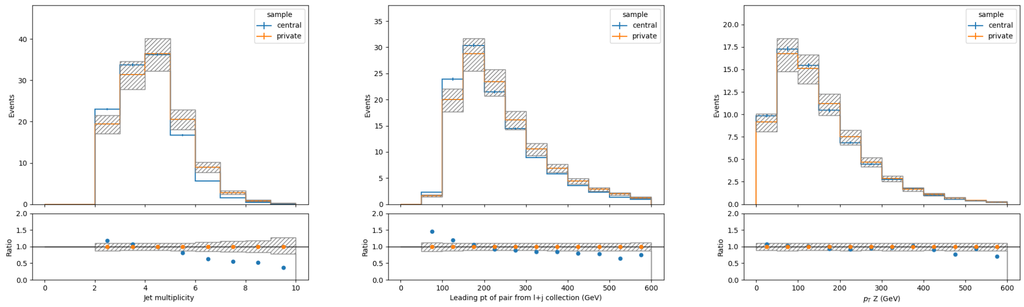
(a) (b) (c)

Figure C.14. RECO level comparison for UL16 $t\bar{t}l\bar{l}$. All other relevant details are the same as described in figure C.1.

change in the shower starting scale) has moved the central UL samples further from our private sample. This makes the tension between our private LO sample and the central NLO sample somewhat worse than was observed in TOP-19-001. However, since there is not reason to believe the change in the shower starting scale represents an improved modeling of the central $t\bar{t}l\bar{l}$ sample, it is not believed that this would imply that the private $t\bar{t}l\bar{l}$ sample is somehow less-well modeled in this analysis than in TOP-19-001. The $t\bar{t}l\bar{l}$ modeling in the private LO samples was already deemed to be acceptable for TOP-19-001. Thus, for these reasons, it is not believed any additional uncertainties would be required in order to account for the differences between the LO and NLO predictions for $t\bar{t}l\bar{l}$.

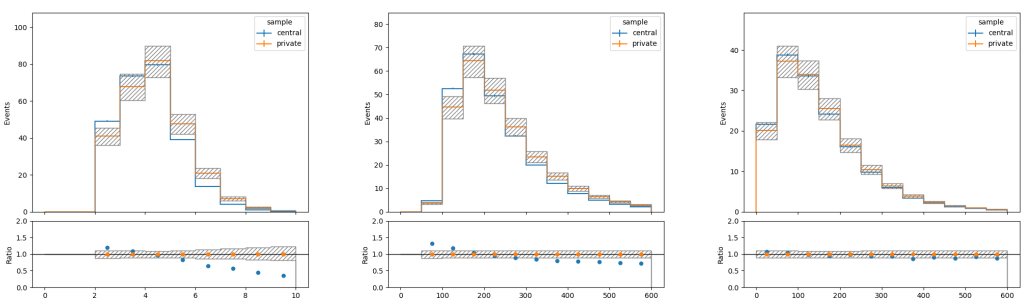
C.5 Summary of comparisons for the $tl\bar{l}q$ sample

As explained in Section 3.2.1, the LO $tl\bar{l}q$ sample (along with tHq) cannot be generated with an extra parton in the matrix element. For this reason, we may expect the modeling to be less accurate (especially at higher jet multiplicities). As in TOP-19-001, we apply an additional systematic uncertainty to this sample in



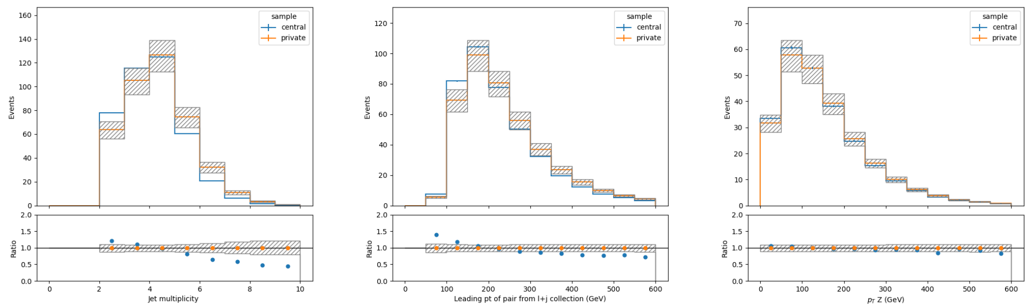
(a) (b) (c)

Figure C.15. RECO level comparison for UL16APV $t\bar{t}l\bar{l}$. All other relevant details are the same as described in figure C.1.



(a) (b) (c)

Figure C.16. RECO level comparison for UL17 $t\bar{t}l\bar{l}$. All other relevant details are the same as described in figure C.1.



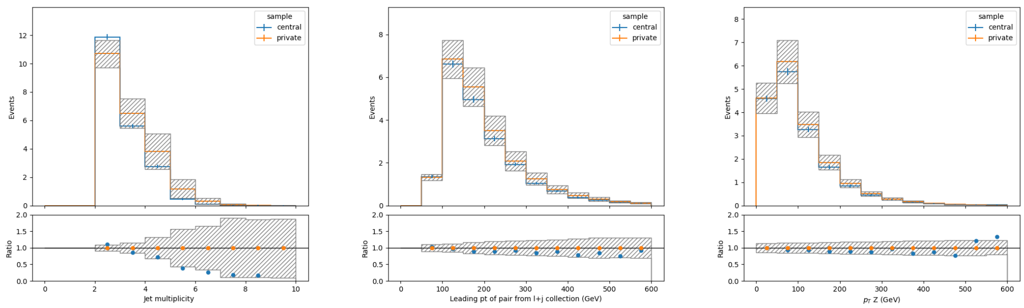
(a) (b) (c)

Figure C.17. RECO level comparison for UL18 $t\bar{t}l\bar{l}$. All other relevant details are the same as described in figure C.1.

order to account for these potential discrepancies. As described in Chapter 9, this uncertainty is derived based based on the comparison between the central NLO $tl\bar{l}q$ N_{jets} distribution, and the private LO $tl\bar{l}q$ N_{jets} distribution. This uncertainty is referred to as the "missing parton" uncertainty. For consistency, the missing parton uncertainty is also applied to the tHq sample (for which we are similarly unable to include an additional parton).

The missing parton systematic was derived and applied in a similar manner in TOP-19-001. However, this analysis takes a more differential approach than TOP-19-001, fitting differential distributions in each jet bin, so it is important to check that this systematic (which was derived in N_{jets}) also covers any discrepancies in the relevant differential distributions used in this analysis. As shown in figures C.18, C.19, C.20, and C.21, the systematic uncertainties (indicated by the shaded grey bands, which includes the "missing parton" uncertainty) generally cover the discrepancies between the samples for the $lj0pt$ and Zp_T distributions (the kinematic distributions used in this analysis).

However, it should be mentioned that the missing parton systematic does not seem

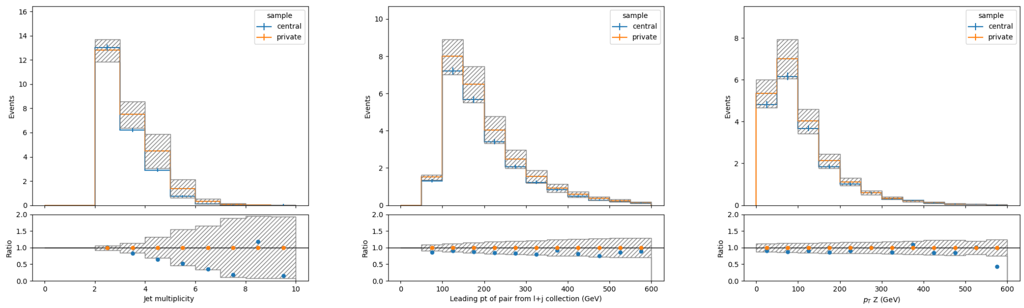


(a) (b) (c)

Figure C.18. RECO level comparison for UL16 $\mathrm{tl\bar{l}q}$. All other relevant details are the same as described in figure C.1.

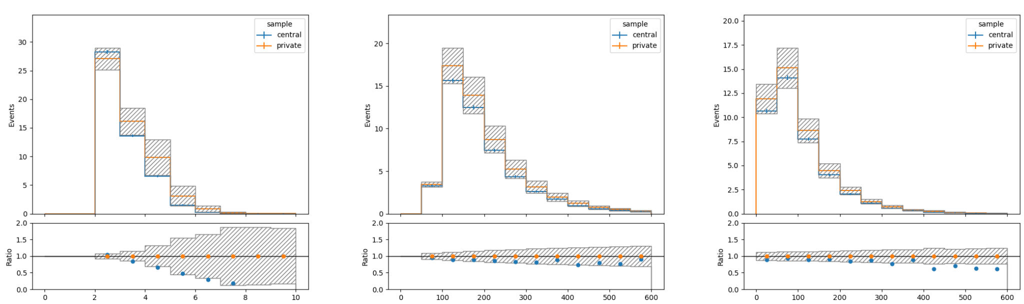
to fully cover the private vs central discrepancies for all kinematic distributions. For example, we have observed that there is tension between the central and private samples for the distribution of the invariant mass of the leading two leptons. The comparison for this distribution for the UL17 samples is shown in figure C.22; the tension is similar for other years as well. Upon additional investigation, we have found that the invariant mass of the samples agrees well when we restrict ourselves to lepton pairs coming from the Z/γ^*, so the discrepancy does not seem to be related to leptons from Z/γ^*.

Nevertheless, it should be emphasized that the invariant mass variable is not one of the variables that is directly used in the analysis. Thus, since the missing parton uncertainty covers the discrepancy for the relevant differential distributions used in the analysis ($lj0pt$ and Zp_T), we believe the modeling of this processes is sufficient, and there would not be a need to implement any additional uncertainties.



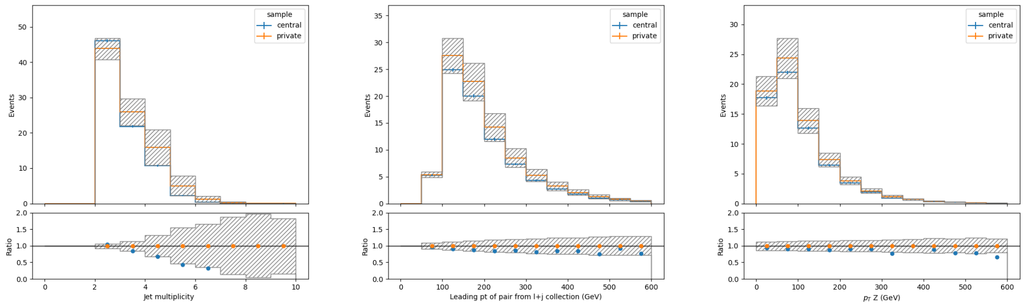
(a) (b) (c)

Figure C.19. RECO level comparison for UL16APV $t l \bar{l} q$. All other relevant details are the same as described in figure C.1.



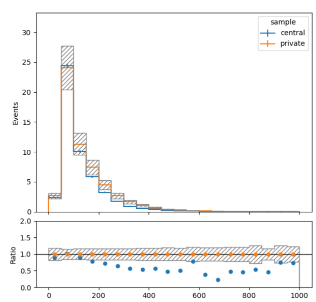
(a) (b) (c)

Figure C.20. RECO level comparison for UL17 $t l \bar{l} q$. All other relevant details are the same as described in figure C.1.



(a) (b) (c)

Figure C.21. RECO level comparison for UL18 $tl\bar{l}q$. All other relevant details are the same as described in figure C.1.



(a)

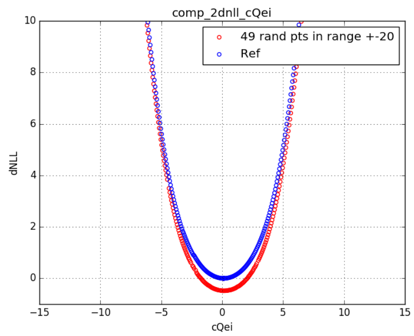
Figure C.22. RECO level comparison for UL17 $tl\bar{l}q$. All other relevant details are the same as described in figure C.18. As discussed in the text, there is tension between the central and private $tl\bar{l}q$ samples for this distribution (invariant mass of the leading two leptons). However, this distribution is not directly used in the analysis.

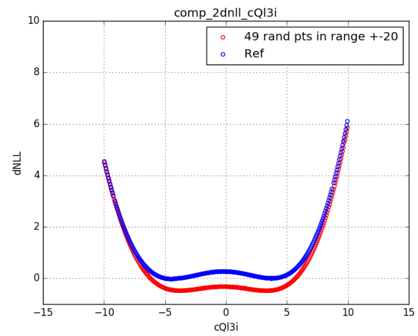
APPENDIX D

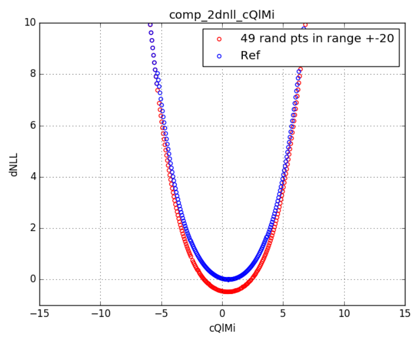
EXAMPLE USAGE OF RANDOM STARTING POINT APPROACH FOR NAVIGATING FALSE MINIMA

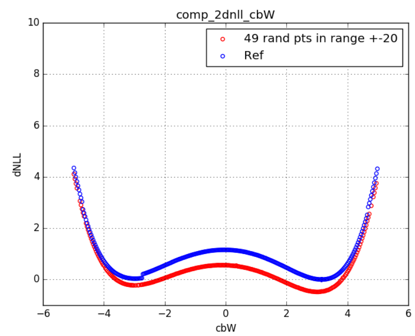
This appendix illustrates how the random starting point method can help to navigate local minima, using the Ref. [18] analysis as an example. The 1d profiled likelihood fits for each of the 16 WCs from [18] are first run without any random starting points; the results of these fits are shown by the blue points in Figure D.1. The blue scan points show several discontinuities, indicating that the fit had become "stuck" in a local minimum. We then performed the scan with 49 random starting points for the profiled parameters, shown in red in Figure D.1. The red scan points show continuous NLL values, and in many cases the random starting point scan was able to identify deeper global minima than had been identified in the original scan.

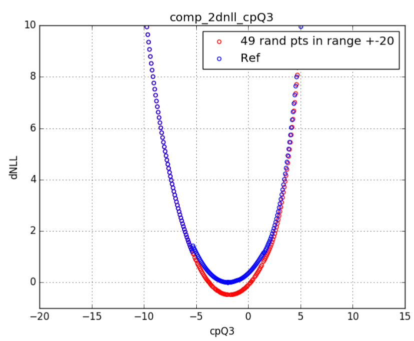
In Figure D.1, the y axis has been scaled such that $y = 0$ corresponds to the NLL obtained from the pre-fit fit that `Combine` performs prior to the scan. In this pre-fit fit, all 16 WCs are profiled. Ideally, this fit should identify the true global minimum. If the pre-fit fit correctly identifies the true global minimum, it would not be possible for a scan point to find a better likelihood value than the one obtained by the pre-fit fit, so there would never be a point on any scan that is below zero. However, in Figure D.1 we see many scanpoints with NLL values below zero (i.e. better than the NLL from the pre-fit fit). This is an indication that the pre-fit for [18] was also challenged by local minima. As discussed in Section 10.3, the [18] analysis worked around this issue by making use of the information in 2d scans. However, the random starting point approach is more general workaround to the issue.

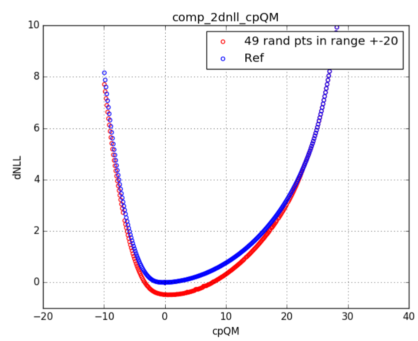

comp_2dnll_cQei
49 rand pts in range +-20
Ref
dNLL
cQei

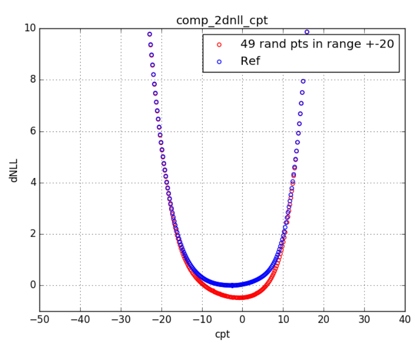

comp_2dnll_cQl3i
49 rand pts in range +-20
Ref
dNLL
cQl3i

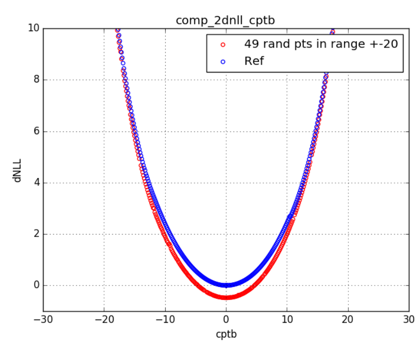

comp_2dnll_cQlMi
49 rand pts in range +-20
Ref
dNLL
cQlMi


comp_2dnll_cbW
49 rand pts in range +-20
Ref
dNLL
cbW


comp_2dnll_cpQ3
49 rand pts in range +-20
Ref
dNLL
cpQ3


comp_2dnll_cpQM
49 rand pts in range +-20
Ref
dNLL
cpQM


comp_2dnll_cpt
49 rand pts in range +-20
Ref
dNLL
cpt


comp_2dnll_cptb
49 rand pts in range +-20
Ref
dNLL
cptb

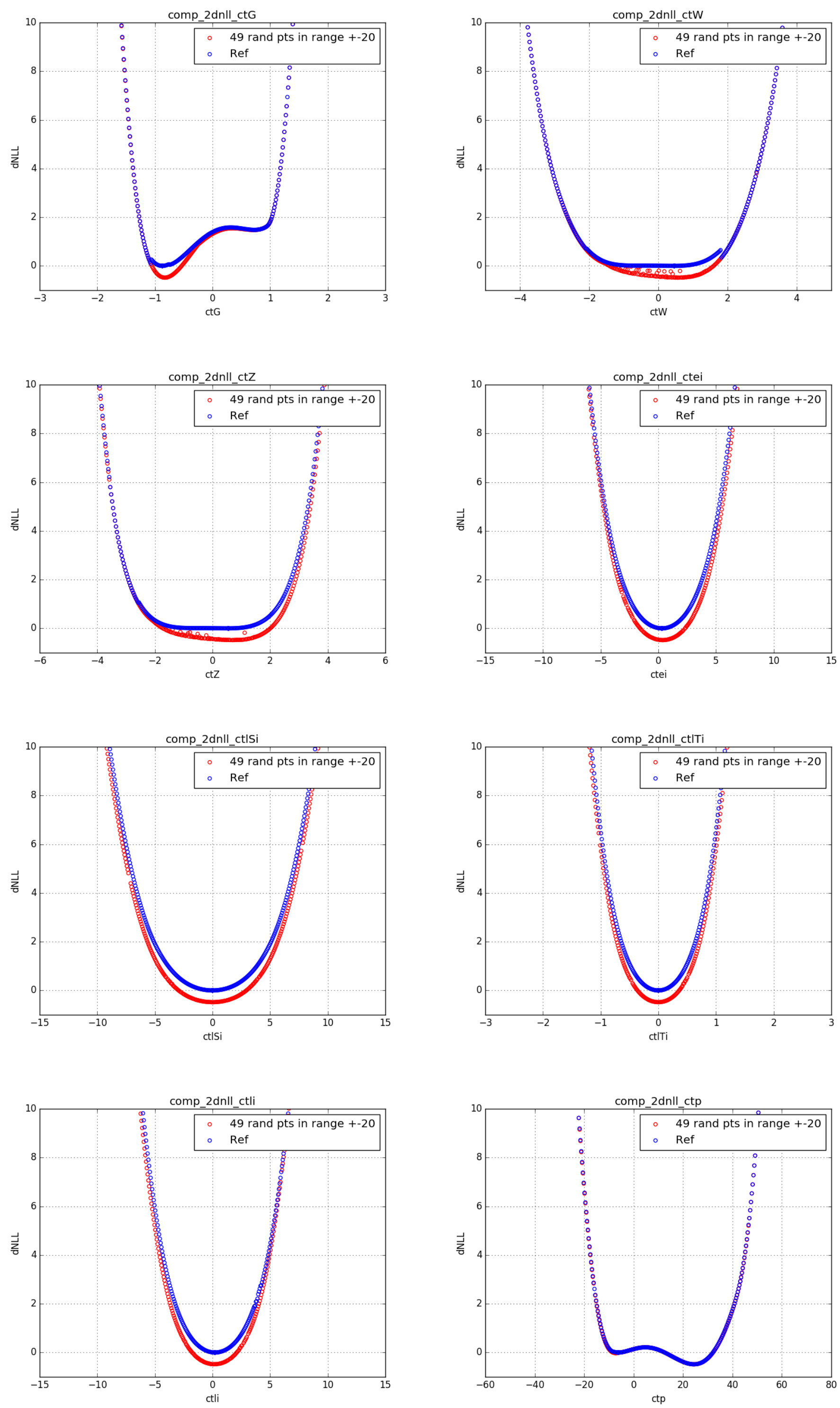


Figure D.1. Profile fits from the [18] analysis with (red) and without (blue) random starting points for the profiled POIs.

BIBLIOGRAPHY

1. 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*, 04:159, 2014. doi: 10.1007/JHEP04(2014)159.

2. J. Alwall, S. de Visscher, and F. Maltoni. QCD radiation in the production of heavy colored particles at the LHC. *JHEP*, 02:017, 2009. doi: 10.1088/1126-6708/2009/02/017.

3. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

4. J. Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008. doi: 10.1140/epjc/s10052-007-0490-5.

5. D. Barducci et al. Interpreting top-quark LHC measurements in the standard-model effective field theory. 2 2018.

6. C. Degrande, N. Greiner, W. Kilian, O. Mattelaer, H. Mebane, T. Stelzer, S. Willenbrock, and C. Zhang. Effective Field Theory: A Modern Approach to Anomalous Couplings. *Annals Phys.*, 335:21–32, 2013. doi: 10.1016/j.aop.2013.04.016.

7. J. L. Feng. Dark matter candidates from particle physics and methods of detection. *Ann. Rev. Astron. Astrophys.*, 48:495, 2010. doi: 10.1146/annurev-astro-082708-101659.

8. R. Goldouzian, J. H. Kim, K. Lannon, A. Martin, K. Mohrman, and A. Wightman. Matching in $pp \to t\bar{t}W/Z/h+$ jet SMEFT studies. *JHEP*, 06:151, 2021. doi: 10.1007/JHEP06(2021)151.

9. B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP*, 10:085, 2010. doi: 10.1007/JHEP10(2010)085.

10. F. James. MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1. 1994.

11. P. Lenzi and J. M. Butterworth. A Study on Matrix Element corrections in inclusive Z/ gamma* production at LHC as implemented in PYTHIA, HERWIG, ALPGEN and SHERPA. 3 2009.

12. Lumi POG. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM.

13. A. Massironi and G. Boldrini. EFT model for SMP measurements. *CMS Analysis Note*, 2020/204, 2020.

14. O. Mattelaer. On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy. *Eur. Phys. J. C*, 76(12):674, 2016. doi: 10.1140/epjc/s10052-016-4533-7.

15. P. J. E. Peebles and B. Ratra. The cosmological constant and dark energy. *Rev. Mod. Phys.*, 75:559, 2003. doi: 10.1103/RevModPhys.75.559.

16. T. A. Porter, R. P. Johnson, and P. W. Graham. Dark matter searches with astroparticle data. *Ann. Rev. Astron. Astrophys.*, 49:155, 2011. doi: 10.1146/annurev-astro-081710-102528.

17. H.-C. project. “HiggsAnalysis-CombineLimit” [software], commit a1dc956452f5b8e055de43bfb093e5243a1e99d9. 2018. URL `https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit`.

18. A. M. Sirunyan et al. Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV using effective field theory. *JHEP*, 03:095, 2021. doi: 10.1007/JHEP03(2021)095.

19. T. Sjostrand, S. Mrenna, and P. Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008. doi: 10.1016/j.cpc.2008.01.036.

20. W. Verkerke and D. P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003.

21. S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60 – 62, 1938. doi: 10.1214/aoms/1177732360. URL `https://doi.org/10.1214/aoms/1177732360`.

22. R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi: 10.1093/ptep/ptac097.

This document was prepared & typeset with pdfLaTeX*, and formatted with* NDdiss2$_\varepsilon$ *classfile (v3.2017.2[2017/05/09])*