

Summer student report

Towards a differential measurement of $t\bar{t}t\bar{t}$ production

Torben Mohr

Duration: 01/07/24 - 27/09/24

Supervisors: Ankita Mehta & Matteo Defranchis

Contents

1	Motivation	1
1.1	Introduction	1
1.2	Current experimental status	1
1.3	Project goal	2
2	First look into the data	5
2.1	MC simulation	5
2.2	The 2ℓ -channel	5
3	Basics of the neural network trainings	8
3.1	Standard NN training	8
3.2	Systematic-aware training	9
4	Results	12
5	Conclusion and outlook	15
	Bibliography	16

1 Motivation

1.1 Introduction

With a predicted cross section of $\sigma = 15.82 \text{ fb}$ at a centre of mass energy $\sqrt{s} = 13.6 \text{ TeV}$ the $t\bar{t}t\bar{t}$ production process is one of the rarest processes currently studied at the Large Hadron Collider (LHC). In the Standard Model (SM) representative diagrams for this process are shown in fig. 1.

Even though it is a rare process, this process has the potential to constrain parameters of the Standard Model Effective Field Theory (SMEFT). In the SMEFT framework the SM Lagrangian is extended by introducing higher dimensional operators built of SM fields while respecting all known SM symmetries. The effective Lagrangian can then be written as

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i, \quad (1.1)$$

where Λ defines the energy scale where the effective field theory approach breaks down and new physics in the form of new particles are expected to arise. The coefficients c_i parametrize how strong an operator \mathcal{O}_i , built from SM fields, contributes to the theory.

The $t\bar{t}t\bar{t}$ process is interesting for indirect searches for new physics phenomena since it is the simplest process to probe top quark self-interactions, parameterized by

$$\mathcal{O}_{tt1} = \frac{c_{tt1}}{\Lambda^2} (\bar{t}\gamma^\mu t)(\bar{t}\gamma_\mu t). \quad (1.2)$$

While this operator only enters through loop processes in other currently accessible processes, it appears at tree level in $t\bar{t}t\bar{t}$ production as shown in fig. 2. When including this operator in the prediction, the number of events in a given phase-space can be written as

$$N \propto \underbrace{|\mathcal{A}_{\text{SM}}|^2}_{\text{SM}} + \underbrace{\frac{c_{tt1}}{\Lambda^2} 2\Re(\mathcal{A}_{\text{SM}} \cdot \mathcal{A}_{\mathcal{O}_{tt}^\dagger}^\dagger)}_{\text{linear}} + \underbrace{\frac{c_{tt1}^2}{\Lambda^4} |\mathcal{A}_{\mathcal{O}_{tt}^\dagger}|^2}_{\text{quadratic}}. \quad (1.3)$$

Due to the contact interaction nature of the four top self-interaction operator, a growth of the EFT contributions as a function of the partonic center-of-mass energy $\sqrt{\hat{s}}$ is expected. Therefore, it is expected that the invariant mass of the four top quarks system $m_{t\bar{t}t\bar{t}}$ will be one of the most sensitive variables, in a sense that it allows to define a region where the quadratic contribution of eq. (1.3) dominates. In fig. 3 this expectation is found to be confirmed. It is also clearly visible that the quadratic term dominates, and that the linear term is especially for higher invariant masses negligible compared to the quadratic contribution.

1.2 Current experimental status

Both the ATLAS and CMS collaborations observed the $t\bar{t}t\bar{t}$ process with Run2 data with a significance of 6.1σ and 5.6σ , respectively [1, 2]. Both observations were done using only the multilepton channels, since they provide the highest (expected) significance as

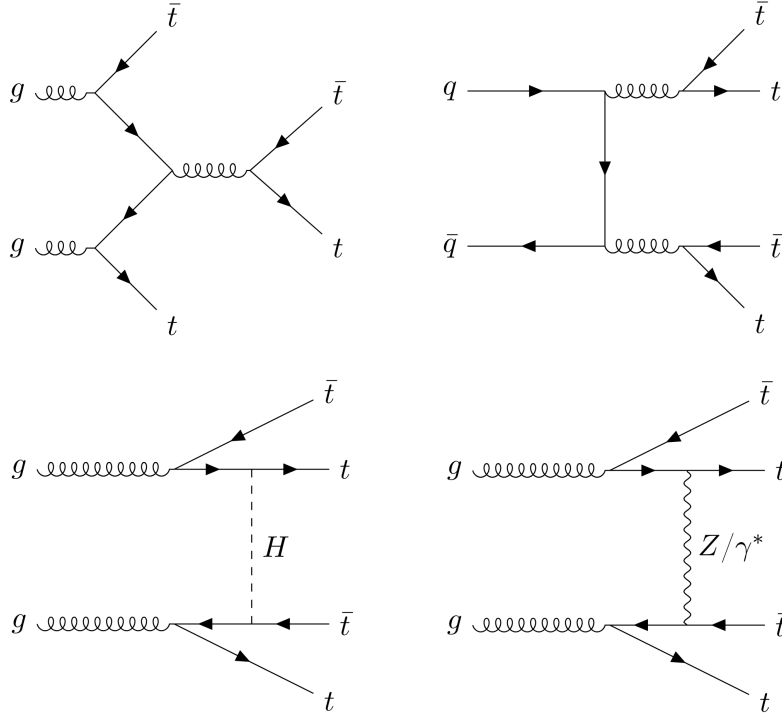


Figure 1: Representative leading order Feynman diagrams for the SM $t\bar{t}t\bar{t}$ production process in pp collisions.

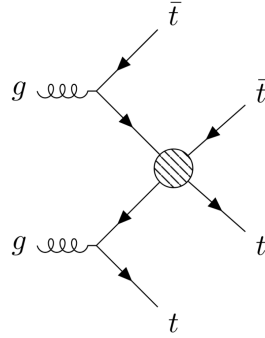


Figure 2: Leading order Feynman diagram where the blob represents the single insertion of the dimension-six EFT operator.

can be seen in fig. 4. The 3ℓ -channel also provides high sensitivity but has a larger background contribution, and will be included at a later stage to further improve the results. Especially the 2ℓ -ss channel provides a high expected significance, which is why the following investigations mainly focus on this final state.

1.3 Project goal

The LHC Run3 is characterized by an increase in center-of-mass with respect to 13 TeV of the LHC Run2 data-taking period to 13.6 TeV which results in a 20% increase in the $t\bar{t}t\bar{t}$ production cross section, while it is expected that the cross section of possible background contributions grow less drastically. Furthermore it is expected to collect data of at least $L_{\text{int}} \approx 300 \text{ fb}^{-1}$. Compared to the CMS observation of this process [1] these two effects should increase the expected number of events by over 250%, which opens up the possibility to perform a differential measurement of the $t\bar{t}t\bar{t}$ process. In the following report it will be shown that the limits that can be set on c_{tt1} strongly contribute from the realization of such a differential measurement.

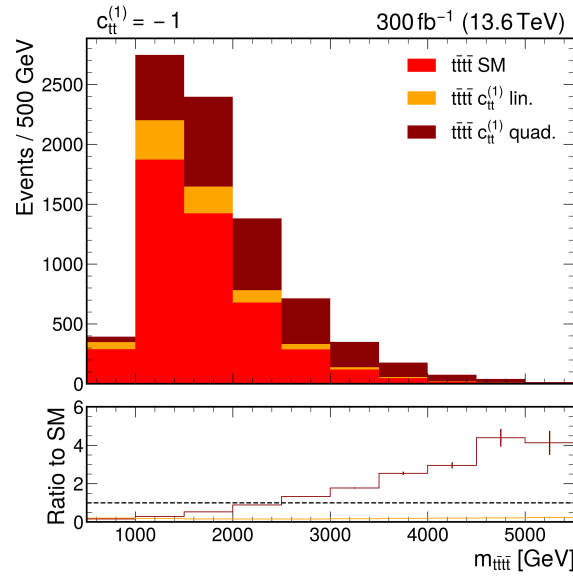


Figure 3: Invariant mass of the four top quark system at generator level. The quadratic contribution from the \mathcal{O}_{tt1} -operator of the EFT dominates at higher masses.

Furthermore the goal is to find a optimal observable that can be used to perform this measurement.

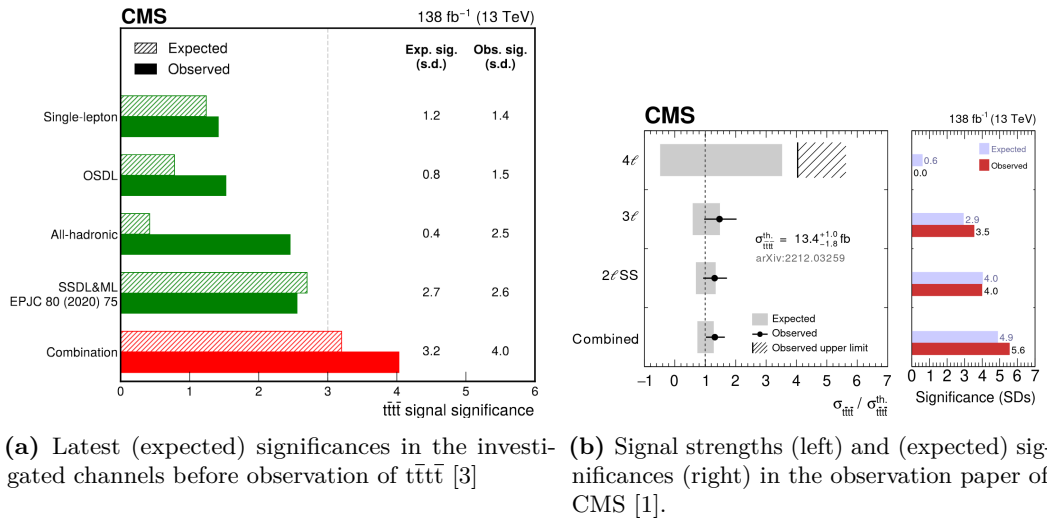


Figure 4: (Expected) significances of the different channels in both the paper claiming evidence (left), and the observation (right) of the $t\bar{t}t\bar{t}$ process.

2 First look into the data

2.1 MC simulation

The samples investigated in this study are purely simulation-based. The Monte Carlo (MC) simulation is done with MadGraph v3.5.1 using the Feynman rules programmed into the SMEFTatNLO UFO model, and using Madspin for the decay of the top quarks and vector bosons involved in these processes. For the $t\bar{t}t\bar{t}$ process, the reweighting method supported by MadGraph is utilized to vary the Wilson coefficients. The hadronization and parton showering is performed with Pythia8.4, and the detector simulation with Delphes3.5.0. In the detector simulation, a setting including PileUp with a mean number of 62 collisions per bunch crossing is used. All samples are currently simulated at leading order, but next-to-leading order contributions in QCD will be included in future development of the project.

2.2 The 2ℓ -channel

The 2ℓ -ss channel is constructed from two leptons with $p_T > 20$ GeV, $|\eta| < 2.4$, and same charge. The same charge selection drastically reduces background contributions from processes involving Z bosons, and can therefore increase the sensitivity. In total this decay channel accounts for approximately 9.7% of all possible decays. However, the expected yield is additionally reduced by the kinematic selections, and by the fact that only leptonically decaying tau leptons are considered in the analysis. The main background in this channel comes from the $t\bar{t}W$ process. By applying a selection on the number of jets this can however be drastically reduced. In fig. 5 the number of jets for both the processes are shown. This distribution only includes jets with $p_T > 30$ GeV, $|\eta| < 2.4$, and additionally the jet collection is cleaned against the two leptons and hadronically decaying tau leptons¹. By applying a cut at $n_{\text{jet}} > 4$ the background is reduced by approximately a factor 14, while having a signal efficiency of 70%. As mentioned in section 1.1, sensitivity to changes introduced by $\mathcal{O}_{tt}^{(1)}$ are expected in energy-related variables. To confirm this the distributions of a selection of variables is shown in fig. 6. The two variables are defined as

$$H_T^{\text{sel.}} = \sum_{\text{sel. jets}} p_T^{j_i}, \quad (2.1)$$

$$\text{MIB} = \left| \sum_{\text{sel. obj.}} \vec{p}_T^i \right|, \quad (2.2)$$

where the first sum runs over all selected jets in the event, and the second sum runs over all selected objects (i.e. the selected jets and the two selected leptons) in the event. The latter variable is defined similar to the Missing Transverse Energy (MET), but since it does not include all activity in the event, the name is changed to *Momentum Imbalance* (MIB). This alternative definition makes MIB only subject to uncertainties like the JES, but not MET-related uncertainties that are expected to be less controllable. The distributions show that the highest sensitivity is observed with hadronic variables, i.e. the leading jet p_T , the

¹Which is done using gen information of the simulation.

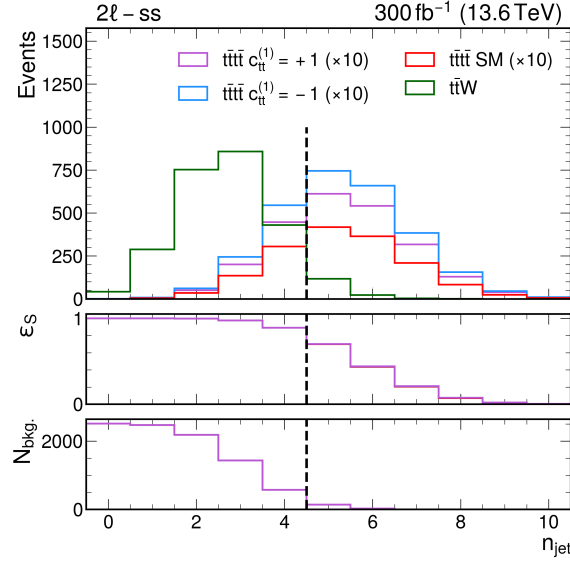


Figure 5: Number of jet distribution for the signal process ($t\bar{t}t\bar{t}$) and the main background ($t\bar{t}W$). The ratio shows that there is no change in signal efficiency when going to EFT (the lines overlap exactly).

sum of all jet p_T , and the momentum imbalance. Compared to this, angular variables like $\Delta\eta(\ell_1, \ell_2)$ show a flat ratio of the quadratic contribution to the SM contribution to the $t\bar{t}t\bar{t}$ process, and therefore not further considered in the studies.

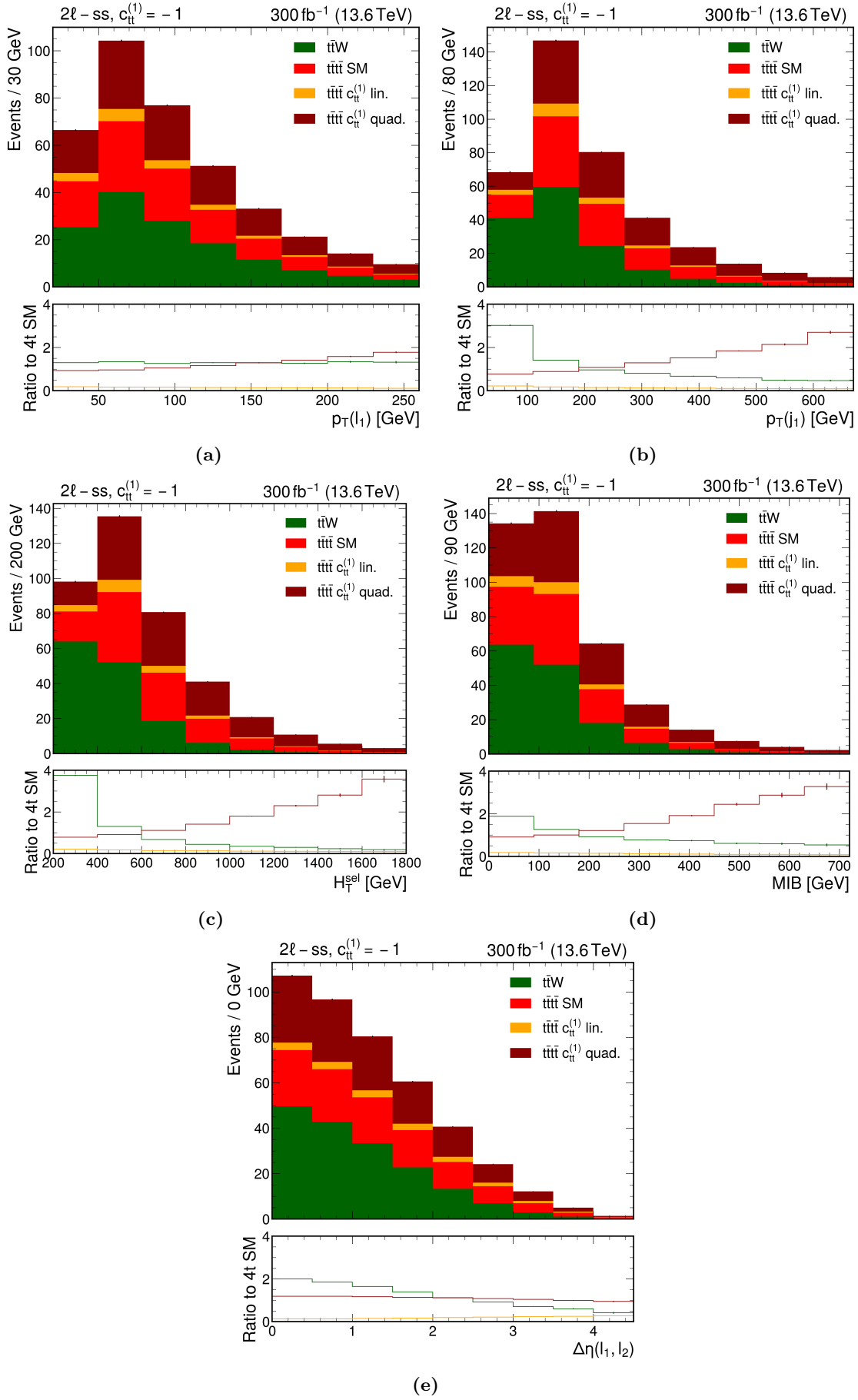


Figure 6: Some example distributions showing in which parts of the phase-space an excess of the EFT contributions can be expected to occur in the 2ℓ -ss channel.

3 Basics of the neural network trainings

The goal of this study is to find an observable that is the mixture of particle-level information that yields the best limit on the Wilson coefficient c_{tt1} . To find this, a feedforward neural network (NN) is used. The input variables are chosen to be

- leading lepton p_T ,
- subleading lepton p_T ,
- leading jet p_T ,
- MIB,
- H_T^{sel} ,
- sum of leading and subleading lepton p_T .

For the training the quadratic contribution to the $t\bar{t}t\bar{t}$ process is defined as signal, whereas the SM contribution (and $t\bar{t}W$) is defined as background.

3.1 Standard NN training

One approach to the NN training is to train it as a standard classifier using the binary cross entropy defined as

$$\text{BCE} = \frac{1}{N} \sum_{i=1}^N c_i \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right], \quad (3.1)$$

where y_i is the label (i.e. 0 for background, 1 for signal) of the event, \hat{y}_i is the NN output, and c_i is the event of the weight, which in this case defined as

$$c_i = \frac{w_i}{\sum_j w_j}, \quad (3.2)$$

where w_i is the corresponding “physics weight”. For SM processes, the physics weight can be defined as

$$w_i = \frac{w_i^{\text{gen}}}{\sum_j w_j^{\text{gen}}} \cdot \mathcal{B} \cdot L_{\text{int}} \cdot \sigma, \quad (3.3)$$

where w_i^{gen} is the gen weight of event i from the Monte Carlo simulation (which is usually constant for LO production), \mathcal{B} is the branching ratio of the corresponding final state, L_{int} is the integrated luminosity, and σ is the cross section of the process. To obtain the weight of the quadratic contribution, reweighting is used:

$$w_i^{\text{quad.}} = \frac{1}{2} \left(w_i(c_{tt1} = +1) + w_i(c_{tt1} = -1) - w_i(c_{tt1} = 0) \right). \quad (3.4)$$

This means that for each event i , three weights are generated. One is the SM weight (indicated by $c_{tt1} = 0$) and then the two weights to reweight according to the EFT scenario of $c_{tt1} = \pm 1$.

The output distribution of the NN should in this case show a separation between the SM contribution of the $t\bar{t}t\bar{t}$ process and the quadratic contribution. It can also be shown that, assuming each event to be binomially distributed (either *signal*, or *background*), the minimization of the BCE loss is equivalent to maximizing the likelihood. Due to the efficiency property of maximum likelihood estimators this means that minimizing the BCE loss is equivalent to minimizing the statistical uncertainty on the signal strength (at least in the large sample limit).

3.2 Systematic-aware training

The second approach incorporates systematic uncertainties during the training according to MLG-23-005. The idea is to construct a binned likelihood from the NN output distribution according to

$$\mathcal{L} = \prod_i^{\text{Bins}} \mathcal{P}(k_i | \lambda_i) \prod_j^{\text{Nuisances}} \mathcal{N}(\tilde{\theta} | \theta) \quad (3.5)$$

where k_i is the observed number of entries in bin i , and

$$\lambda_i = n_i^{\text{t}\bar{t}\text{t}\bar{t}, \text{SM}} + n_i^{\text{t}\bar{t}\text{W}} + r_s \cdot n_i^{\text{t}\bar{t}\text{t}\bar{t}, \text{quad}} + \Delta_i, \quad (3.6)$$

is the expected number of events in bin i , which are combined in the Poisson distribution \mathcal{P} . In the definition of λ_i in turn the last term describes the shifts induced by all nuisance parameters $\{\theta_j\}$ and all processes:

$$\Delta_i = \sum_{p \in \{\text{t}\bar{t}\text{t}\bar{t}_{\text{SM}}, \text{t}\bar{t}\text{t}\bar{t}_{\text{quad}}, \text{t}\bar{t}\text{W}\}} \sum_j^{\text{Nuisances}} \max(0, \theta_{j_p}) \left(\Delta_{j_p}^{\text{up}} \right)_i + \min(\theta_{j_p}, 0) \left(\Delta_{j_p}^{\text{down}} \right)_i. \quad (3.7)$$

From this likelihood, the uncertainty on r_s can be estimated by setting all nuisances to 0, and using an Asimov dataset. Then the uncertainty on r_s can be computed using the first element of the inverse of the Hessian of the likelihood, i.e.

$$\Delta r_s = \sqrt{F_{r_s, r_s}^{-1}} \quad \text{with} \quad F_{x_i x_j} = - \frac{\partial^2 \log \mathcal{L}}{\partial x_i \partial x_j}_{x_i x_j \in \{\{r_s\}, \{\theta_j\}\}}. \quad (3.8)$$

The uncertainties taken into account during the training are JES variations, and Parton Shower (PS) variations which are expected to have the largest impact on the variables that are used in the training. The JES uncertainty is parametrized according to DP2020-019 and is shown in fig. 7 as a function of the jet p_T . The impact of this uncertainty on the three processes used during the training is shown in fig. 8. In the ratios it can be seen that the statistical fluctuations for the $t\bar{t}t\bar{t}$ quad. and the $t\bar{t}\text{W}$ process are still big, which can affect the performance of the training and lead to less precise or sub-optimal results. This problem can be solved by increasing the statistics of the MC sample.

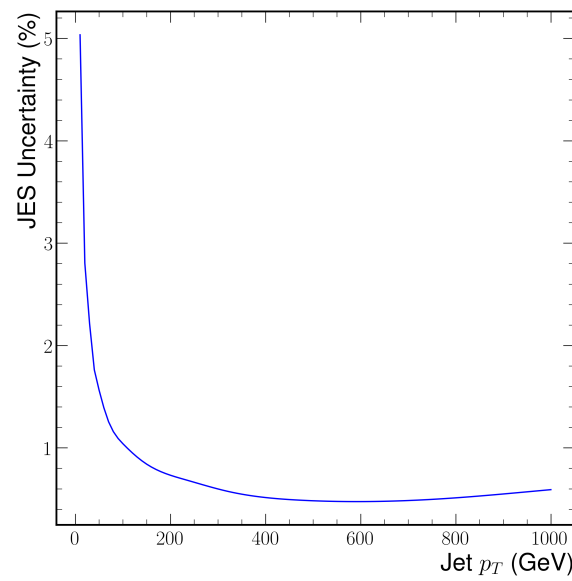


Figure 7: Parametrization of the JES uncertainty.

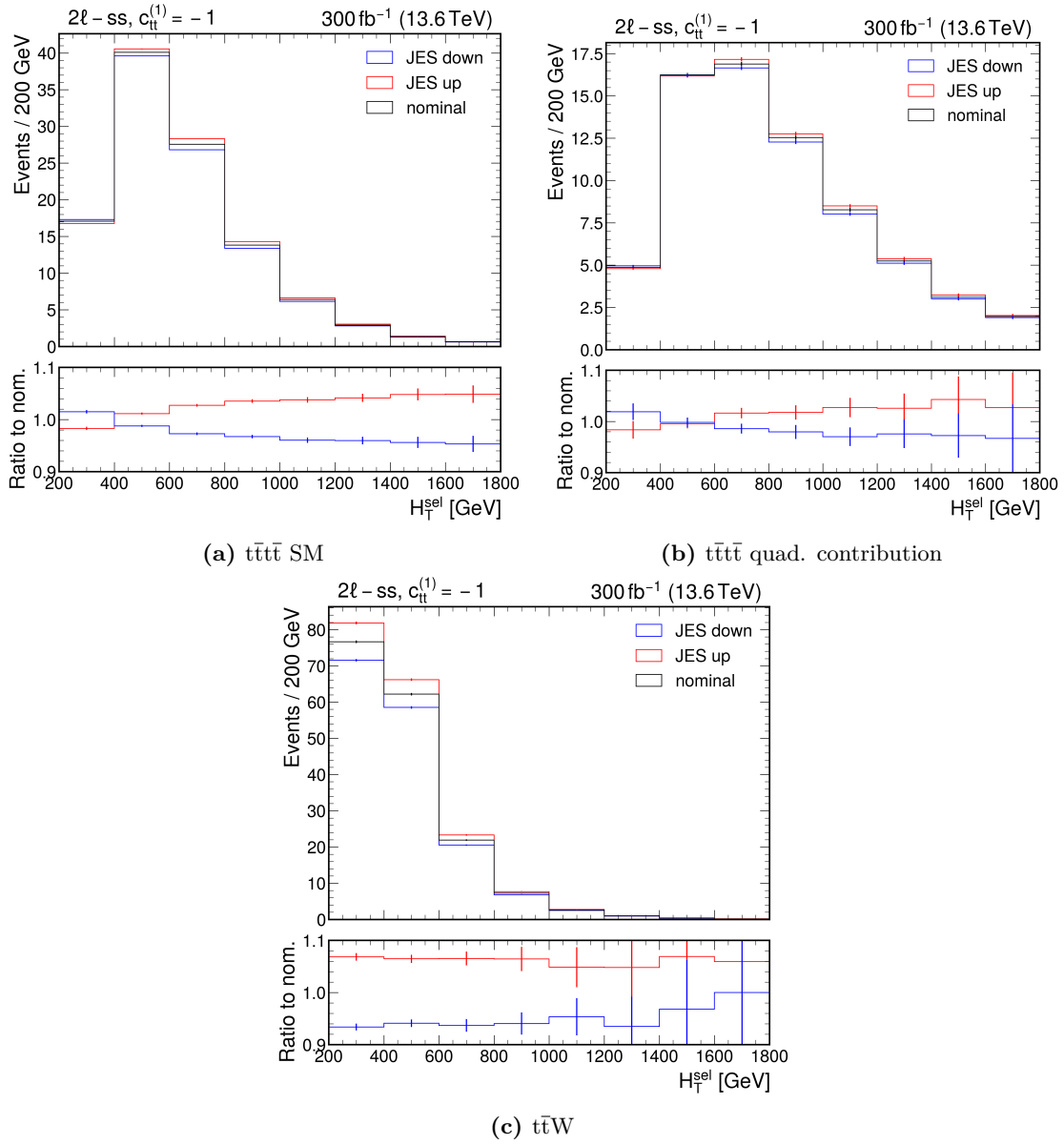


Figure 8: Some example distributions showing in which parts of the phase-space an excess of the EFT contributions can be expected to occur in the 2ℓ -ss channel.

4 Results

First of all the goal was to investigate the the gain of using the systematic aware-training over the normal cross-entropy training. For this, both NN trainings are performed using a dataset of ~ 1.3 billion selected 2ℓ events from the $t\bar{t}t\bar{t}$ process and ~ 40.000 selected 2ℓ events from the $t\bar{t}W$ process (statistics will be improved in the future). The $t\bar{t}t\bar{t}$ events are used twice in the training. Once they're given with the SM weight and classified as background, and once they're given with the quadratic weight and classified as signal.

To compare the two different training approaches, first one can inspect the differences in the output distributions. This comparison is shown in fig. 9, where additionally to the

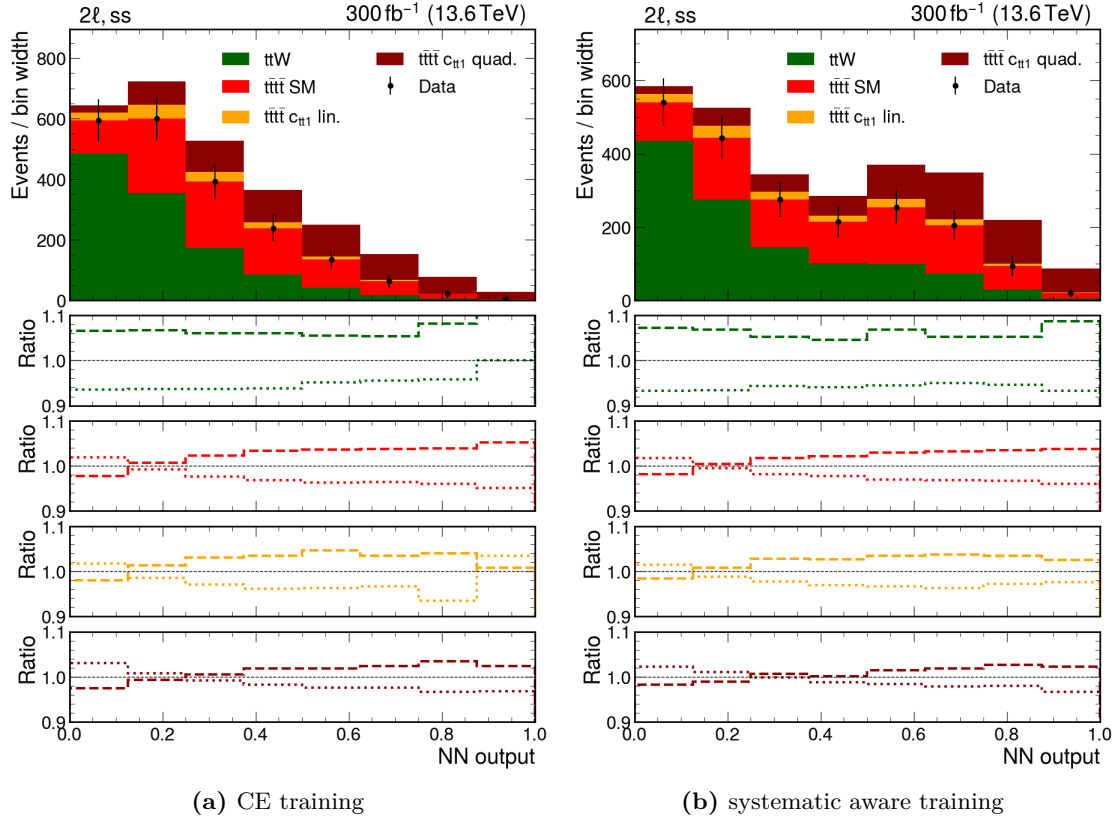


Figure 9: Distributions of the NN output after the training.

nominal distributions, also the up- and down-variations of the JES uncertainty are shown as ratios to the nominal distributions separately for each process. There the tendency of smaller variations for the systematic aware training is visible. However, the effect is pretty small and minor compared to the effect of the statistical uncertainties which are shown for the point labeled “Data”. This indicates the Asimov data that is used for the final fit to extract the expected limits on c_{tt1} .

To measure the performance of the NNs, a fit using COMBINE and the AnalyticAnomalousCoupling-framework is performed. The resulting likelihood scan fig. 10a shows, that the limits

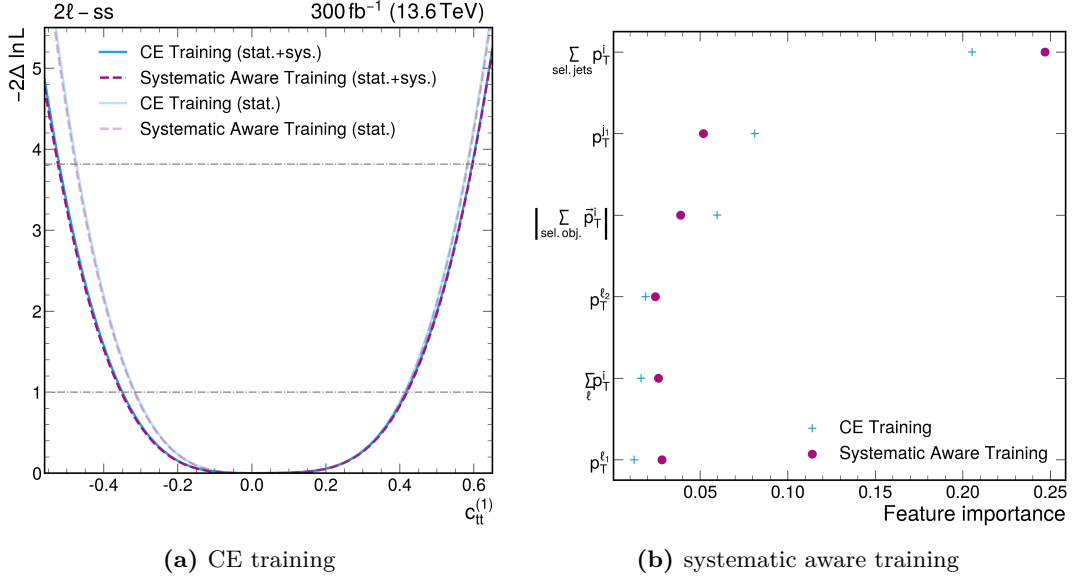


Figure 10: Comparison of the two NNs using a likelihood scan (left), and the feature importance (right).

obtained by both NNs are compatible. The reason for that is probably that the process is statistically dominated as can be seen from the same plot where also the statistical-only likelihood scans are indicated. In the case where the statistical uncertainty dominates, the minimization of the total uncertainty of the signal strength is equivalent to minimizing the statistical uncertainty, which in turn is equivalent to minimizing the BCE loss. Also when comparing the feature importances in fig. 10b it is visible that the systematic aware training does not improve the limits significantly given the current statistics of MC simulations.

To conclude this part, it can be said that with the currently included systematic uncertainties and integrated luminosity of 300 fb^{-1} , the systematic aware training does not bring improvement over the classical CE training given the current statistics of signal simulation.

Finally, a comparison of different ways to obtain a limit on c_{tt1} are compared. The most simple way would be to perform an inclusive cross section measurement, which is equivalent to performing a fit with a single bin. In fig. 11 it can be clearly seen, that adding any differential information (i.e. using more than one bin) narrows the likelihood in any case. When e.g. performing a differential measurement with three bins in the NN output, the lower (upper) bound of the 95% CI can be set 34% (23%) more stringent compared to the inclusive cross section (single bin) measurement. It can also be seen that the difference between performing a measurement with three or six bins in the NN output is contained. The comparison with the ATLAS result [2] is not entirely equitable at this stage, as not all systematic uncertainties have been accounted for. Additionally, the ATLAS result is based on Run 2 data, whereas the findings of this study represent a forecast for Run 3 results. Therefore, this limit should be regarded as a reference for the current experimental limit.

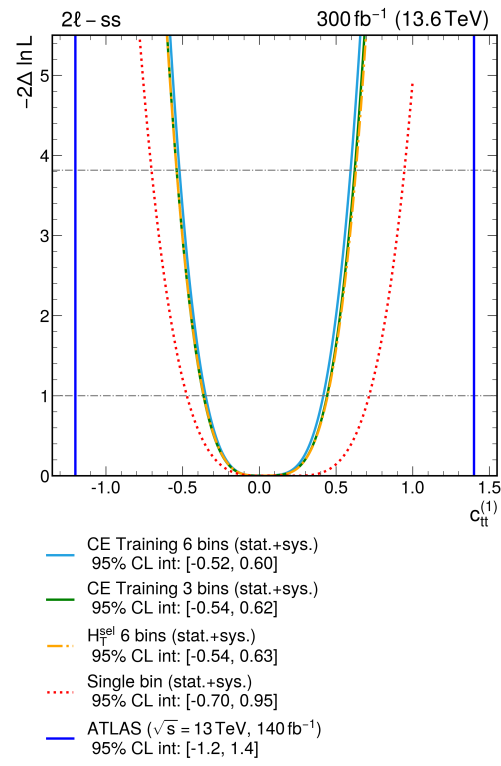


Figure 11: Comparison of different ways to obtain a limit on c_{tt1} .

5 Conclusion and outlook

It was demonstrated that the systematic aware training proposed in MLG-23-005 is in general applicable to this problem. Due to the limited sample statistics it was also concluded that the difference to the classical CE training is negligible. This training method should however be considered for cases with higher integrated luminosity, and cases where more systematics are taken into account.

The final result also clearly showed the significant impact that a differential measurement can have on constraining the Wilson coefficient of the top self-interaction. It was also demonstrated that in the realistic scenario of measuring the differential cross section in three bins using the Run3 dataset, this already improves a lot over a inclusive cross section measurement.

The next steps should be to solve the problems of the statistical fluctuations in the data by further increasing the MC statistics, or by generating a dedicated sample for the “EFT” case, where high-energetic bins are more populated, leading to overall smaller weights, and therefore to smaller statistical uncertainties. After this is fixed, the study should be extended to the 3ℓ channel to further increase the sensitivity.

The final goal will then be to derive an analytical expression of the optimal observable by performing a Symbolic Regression of the neural network output. This expression can then be used to perform the real measurement, and can also be helpful to easily compare to theoretical predictions.

Bibliography

- [1] CMS Collaboration. Observation of four top quark production in proton-proton collisions at $s=13\text{TeV}$. *Physics Letters B*, 847:138290, 2023. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2023.138290>. URL <https://www.sciencedirect.com/science/article/pii/S037026932300624X>.
- [2] ATLAS collaboration. Observation of four-top-quark production in the multilepton final state with the atlas detector. *The European Physical Journal C*, 83(6), June 2023. ISSN 1434-6052. doi: 10.1140/epjc/s10052-023-11573-0. URL <http://dx.doi.org/10.1140/epjc/s10052-023-11573-0>.
- [3] CMS Collaboration. Evidence for four-top quark production in proton-proton collisions at 13 TeV. *Physics Letters B*, 844:138076, September 2023. ISSN 0370-2693. doi: 10.1016/j.physletb.2023.138076. URL <http://dx.doi.org/10.1016/j.physletb.2023.138076>.