Using associated top quark production to probe for new physics within the framework of effective field theory

TOP-19-001
Joker Talk
Brent R. Yates
The Ohio State University
On behalf of the CMS Collaboration
brent.yates@cern.ch
Motivation for new physics

The Standard Model (SM) is wonderfully precise, but only accounts for 5% of the universe. 

Shortcomings include:

- **Dark matter/energy** – Invisible particles? Non-zero vacuum energy?
- **Hierarchy problem** – Why is the Higgs mass $\mathcal{O}(10^2)$ GeV but Plank mass $\mathcal{O}(10^{19})$ GeV?
- **Baryon asymmetry** – Why do we live in a universe devoid of anti-matter?

After discovering the Higgs boson in 2012, the LHC provided no definitive evidence of anything unexpected.

Assume $\Lambda_{NP} > \Lambda_{LHC}$

How might it appear at the LHC?
Introduction to effective field theory

New physics at scales beyond what the LHC can directly probe can be approximated by expanding terms of higher dimensional \((d)\) operators \(\mathcal{O}\) consisting of SM fields

Operators are suppressed by powers of the energy scale \(\Lambda\), and the strength is controlled by the Wilson coefficients (WCs) \(c_i\)

\[
\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_{d,i} \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}
\]

Dimension five violates lepton number

Dimension six is the focus of this analysis

Higher dimensions are suppressed by additional powers of \(\Lambda\)
Analysis overview

A novel technique to examine data collected in 2017
Performs a global fit across all processes (signal and background)

Probe EFT effects using multiple lepton final states

Procedure helps constrain systematic uncertainties

Correlations rely on data (no assumptions made)

Using channels with $t\bar{t}l\nu$, $t\bar{t}l\bar{l}$, $t\bar{t}q$, $t\bar{t}H$, and $tHq$ production ($H \rightarrow W^+W^-$, $ZZ$, $\tau^+\tau^-$, exclude $H \rightarrow b\bar{b}$)
EFT parameterization

Matrix elements $\mathcal{M}$ split into SM and EFT terms

Parameterize cross section by WCs:

- SM terms ($s_{0i}$)
- Interference terms between the SM and EFT ($s_{1ij}$)
- Pure EFT terms ($s_{2ij}$)
- Interference terms between EFT ($s_{3ijk}$)

Individual events ($d\sigma$) have weight $w_i \rightarrow$ can be summed to produce the predicted event yields

\[
\mathcal{M} = \mathcal{M}_{\text{SM}} + \sum_{j} \frac{c_j}{\Lambda^2} \mathcal{M}_j
\]

\[
\sigma \propto \mathcal{M}^2
\]

\[
w_i(\mathcal{C}) = s_{0i} + \sum_{j} \frac{s_{1ij}}{\Lambda^2} + \sum_{j} \frac{s_{2ij} c_j^2}{\Lambda^4} + \sum_{j,k} \frac{s_{3ijk} c_j c_k}{\Lambda^2 \Lambda^2}
\]
Weight for each event: \( w_i(\vec{c}) = f_{SM \ i} + g_{EFT \ i}(\vec{c}) \)

Coefficients summed as new events are added

\[
f_{SM \ i} = s_{0i} \\
g_{EFT}(\vec{c})_i = \sum_j s_{1ij} \frac{c_j}{\Lambda^2} + \sum_j s_{2ij} \frac{c_j^2}{\Lambda^4} + \sum_{jk} s_{3ijk} \frac{c_j c_k}{\Lambda^2} \]
EFT parameterization

MC simulations are generated with non-zero WCs
Extra partons are added when possible

Initial values chosen to include all relevant phase space and optimize the MC statistical power

\[ \sigma_{\text{stat}}^2 = \sum w_i^2(c) \]

Weight of each event accounts for variation in yield from EFT

Used to solve for the constant terms in the quadratic fit

This parameterization will be used in the fit
Dim6TopEFT Model

EFT simulations are generated by MADGRAPH_aMC@NLO using the dim6TopEFT[1] model

- **Warsaw basis** of dimension six operators
- $\Lambda = 1$ TeV
- CKM matrix is assumed to be a unit matrix
- $u, d, s, c, e, \mu$ masses all set to zero
- The unitary gauge is used and Goldstone bosons are removed
- Baryon and lepton number violating operators are not included
- Only tree-level simulation is possible
- Lepton universality is assumed (all flavors set to same WCs)

The 16 operators which have the largest impact on the signal processes, and relatively small impact on $t\bar{t}$ background, are considered

Model operators

Only the real components are considered since the imaginary coefficients lead to CP violation, and are well constrained by EDM experiments and $B \to X_s \gamma$ decays.

Examples of diagrams involving vertices arising from the $O_{tW}$ operator.

Examples of diagrams involving vertices arising from the $O_{t\ell}$ operator.
Event selection

The analysis is split into lepton (ℓ) categories as well as jet multiplicity (both light and b-tagged jets)

A BDT is applied to separate prompt from non-prompt leptons

Final-state observables are an admixture of processes (the method does not require we separate processes)

• Each analysis bin stores the sum of the quadratic coefficients → event yields are fully parametrized by the WCs
Event categorization

<table>
<thead>
<tr>
<th>Jets</th>
<th>2$\ell$ same sign</th>
<th>3$\ell$</th>
<th>$\geq 4\ell$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Sigma_\ell q &lt; 0, \Sigma_\ell q &gt; 0$</td>
<td>$\Sigma_\ell q &lt; 0, \Sigma_\ell q &gt; 0$</td>
<td>$\geq 4\ell$</td>
</tr>
<tr>
<td>4 jets</td>
<td>Z boson</td>
<td>Z boson</td>
<td>Z boson</td>
</tr>
<tr>
<td>5 jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 7$ jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 2$ b-jet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 medium + loose or medium)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sum_\ell q &lt; 0, \sum_\ell q &gt; 0$, $\sum_\ell q &lt; 0, \sum_\ell q &gt; 0$</td>
<td>$\sum_\ell q &lt; 0, \sum_\ell q &gt; 0$</td>
<td>$\geq 4\ell$</td>
<td></td>
</tr>
<tr>
<td>2 jets</td>
<td>Z $\rightarrow$ e$^+$e$^-$</td>
<td>Z $\rightarrow$ e$^+$e$^-$</td>
<td>Z $\rightarrow$ e$^+$e$^-$</td>
</tr>
<tr>
<td>3 jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 5$ jets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 b-jet medium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 2$ b-jet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 medium + loose or medium)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

loose = 85% efficiency, 10% light/quark jets
medium = 70% efficiency, 1% light/quark jets
Probability of a non-prompt lepton passing prompt cuts is measured in a multijet enriched region.

Data-driven
Fitting procedure

Each bin is treated as a Poisson experiment with a probability of obtaining the observed data.

Profiled likelihood simultaneously fits all bin and extract the $2\sigma$ confidence intervals of the various WCs.

Two fitting procedures are used:

Scan single WC, other 15 are unconstrained nuisance parameters
  • More physical of the two, no reason for new physics to only favor one WC

Scan A single WC, other 15 are fixed to their SM value of zero
  • Extreme scenario where nature has a single WC; the ability to fit one is limited by the lack of knowledge of 15 others

18 September 2020
Event yields in lepton multiplicity bins (jet bins integrated)

SM only
(all WCs set to 0)
Event yields in lepton multiplicity bins (jet bins integrated)

EFT reduces predicted yields (WCs = 1/6 final value)
Event yields in lepton multiplicity bins (jet bins integrated)

EFT reduces predicted yields (WCs $= 2/6$ final value)
Event yields in lepton multiplicity bins (jet bins integrated)

EFT reduces predicted yields (WCs = 3/6 final value)
Event yields in lepton multiplicity bins (jet bins integrated)

EFT reduces predicted yields (WCs = 4/6 final value)
Event yields in lepton multiplicity bins (jet bins integrated)

EFT enhances predicted yields (WCs = 5/6 final value)
Event yields in lepton multiplicity bins (jet bins integrated)

WCs set to final values

18 September 2020

Top 2020 Joker Talk
Single WC scan

Scanning $c_t^{S(\ell)}$ while treating the other 15 as unconstrained nuisance parameters or fixed to the SM value of zero.

Degenerate minima widen the NLL curve.
Two-dimensional WC scans

Pairs of WCs are also scanned to help investigate the correlations between WCs, as visualizing the full 16-dimensional hypersurface is not feasible.

Other 14 WCs treated as unconstrained nuisance parameters

Other 14 WCs are fixed to zero (SM)

Better constraints along the negative vertical axis

Looser constraints along the positive vertical axis

Profiled includes the SM

Fixed excludes the SM at 1σ

2 WCs must compensate for other 14 $\rightarrow$ less correlation
Important systematic uncertainties

Analysis specific

- **Misidentified lepton rate estimate** – Contamination from non-prompt leptons
  Overcome by examining the analysis side-bands
  Data-driven → statistically limited

Monte Carlo simulation modeling

- **Matrix element parton shower matching** – Matching extra partons to final-state jets
- **Missing parton uncertainty** – Extra partons cannot be added for $tHq$ and $tllq$
  Compare LO EFT without extra partons to NLO SM simulations, assign uncertainty to cover any discrepancies
  These issues will not be present in SMEFT@NLO, and we are very interested in the development
- **Scale uncertainties** – FSR and ISR
Wilson coefficient CIs

The $1\sigma$ and $2\sigma$ CIs are given

When the other WCs are fixed to zero, the fit can produce degenerate minima in $c_{tW}$, $c_{t\phi}$, $c_{tG}$, and $c_{\phi t}$

Degenerate minima are due to the quadratic nature of the parameterization

None of the WCs exclude the SM point of zero by a statistically significant amount
Conclusion

The production of or more $t$ quarks in association with additional leptons were used to measure the confidence intervals of 16 dimension-six EFT operators using data collected in 2017.

The EFT yields are parameterized using a quadratic function of event weights.

This novel technique allows us to extract EFT from difficult data.

The 2$\sigma$ CIs were extracted for these operators. Intervals are compatible with the SM and other analyses [1].

With the full Run II data set (almost triple the integrated luminosity) more sophisticated analyses may be performed, including using differential distributions.

[1] https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures
Backup
Selecting operators of interest

There are 59 total dimension-6 operators that conserve baryon and lepton number

We only consider 16 operators:

- Operators must appear in signal processes (top + boson) at tree level
- Ignore operators that strongly affect background processes
- Imaginary parts of non-Hermitian operators are set to zero

‘Two heavy + boson’: $c_{t\varphi}$, $c_{\varphi Q}$, $c_{\varphi Q}^3$, $c_{\varphi t}$, $c_{\varphi tb}$, $c_{tW}$, $c_{tZ}$, $c_{bW}$, and $c_{tG}$

‘Two heavy + two lepton’: $c_{Ql}^{3(l)}$, $c_{ql}^{-(l)}$, $c_{Qe}$, $c_{tl}$, $c_{te}$, $c_{t}^{S(l)}$, and $c_{t}^{T(l)}$

- These operators have three copies that couple to the different lepton flavors
- We assume equal coupling to all flavors to reduce the number of operators to these seven
Lepton identification

\[ I_\ell = \sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutral}} p_T - \rho A \left(\frac{R}{0.3}\right)^2\right) \]

\[ R = \begin{cases} 
0.05 & \text{if } p_T > 200 \text{ GeV} \\
10 \text{ GeV}/p_T & \text{if } 50 < p_T < 200 \text{ GeV} \\
0.20 & \text{if } p_T < 50 \text{ GeV} 
\end{cases} \]

<table>
<thead>
<tr>
<th>Electrons</th>
<th>Pseudorapidity range</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 &lt;</td>
<td>(\eta</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>1.0 &lt;</td>
<td>(\eta</td>
<td>&lt; 1.479</td>
</tr>
<tr>
<td>1.479 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.0</td>
</tr>
<tr>
<td>2.0 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.2</td>
</tr>
<tr>
<td>2.2 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.3</td>
</tr>
<tr>
<td>2.3 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.4</td>
</tr>
<tr>
<td>2.4 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Muons</th>
<th>Pseudorapidity range</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 &lt;</td>
<td>(\eta</td>
<td>&lt; 0.8</td>
</tr>
<tr>
<td>0.8 &lt;</td>
<td>(\eta</td>
<td>&lt; 1.3</td>
</tr>
<tr>
<td>1.3 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.0</td>
</tr>
<tr>
<td>2.0 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.2</td>
</tr>
<tr>
<td>2.2 &lt;</td>
<td>(\eta</td>
<td>&lt; 2.5</td>
</tr>
</tbody>
</table>
Inputs to BDT

- Lepton $p_T$ and $\eta$
- $I_{\ell}^{\text{charged}}$
- $I_{\ell}^{\text{neutral}}$
- $p_T^{\ell}/p_T^{\text{jet}}$
- CSVv2 b-tagging algorithm
- $N_{\text{charged}}$ of charged particles within the jet
- $p_T^{\text{rel}} = p_\ell \sin(\theta)$
- $d_{xy}$ and $d_z$ w.r.t. the PV
- $d/\sigma_d$ signed 3D impact parameter significance w.r.t the PV
- Lepton MVA from EGamma POG
- Compatibility of track segments with the muon system
### Event selection

<table>
<thead>
<tr>
<th>Tight Leptons</th>
<th>2/4 ss</th>
<th>2/4</th>
<th>3μ</th>
<th>≥4μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeepCSV(b) Medium Jets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeepCSV(b) Loose Jets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nJet Subcategories</td>
<td>4, 5, 6, ≥7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepton Charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sign of Net Lepton Charge</td>
<td></td>
<td></td>
<td>≥1μ</td>
<td></td>
</tr>
<tr>
<td>“SFOS Z”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Subcategories</th>
<th>2/4/5 ε</th>
<th>2/4(≥2μ)</th>
<th>2/4 (≥2μ)</th>
<th>3(≥2μ)</th>
<th>3(≥2μ)</th>
<th>3(≥2μ)</th>
<th>≥4μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diboson</td>
<td>1.4</td>
<td>1.2</td>
<td>5.9</td>
<td>4.7</td>
<td>0.4</td>
<td>0.3</td>
<td>52.1</td>
</tr>
<tr>
<td>Triboson</td>
<td>1.0</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Charge Flips</td>
<td>1.5</td>
<td>1.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Fakes</td>
<td>25.6</td>
<td>26.4</td>
<td>11.3</td>
<td>13.0</td>
<td>3.3</td>
<td>2.5</td>
<td>16.9</td>
</tr>
<tr>
<td>Conversions</td>
<td>19.0</td>
<td>19.2</td>
<td>2.3</td>
<td>2.6</td>
<td>1.7</td>
<td>1.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Sum Background</td>
<td>47.0</td>
<td>46.2</td>
<td>19.7</td>
<td>20.5</td>
<td>5.4</td>
<td>4.8</td>
<td>73.3</td>
</tr>
</tbody>
</table>

| ttV                | 61.2    | 37.1     | 15.4     | 8.0    | 10.8   | 5.9    | 2.9  |
| ttZ                | 19.3    | 19.0     | 12.7     | 13.3   | 9.1    | 8.5    | 955.5|
| tth                | 28.7    | 24.1     | 7.9      | 7.6    | 5.1    | 5.2    | 3.2  |
| thq                | 2.7     | 3.5      | 3.5      | 1.8    | 1.2    | 0.6    | 59.8 |
| thq                | 0.8     | 0.4      | 0.3      | 0.2    | 0.2    | 0.2    | 0.1  |

| Sum Signal         | 116.2   | 82.2     | 58.9     | 50.8   | 26.3   | 20.3   | 141.6|

| Total Expected     | 163 ± 20| 128 ± 15 | 59 ± 7  | 51 ± 6 | 32 ± 4 | 25 ± 3 | 215 ± 25|
| Data               | 192     | 171      | 85      | 64     | 32     | 26     | 239   |
Background Estimation

Misidentified leptons

- **Data-driven** by dividing data into measurement and application regions (analysis side-bands)
- Measurement region contains QCD multijet background dominated by non-prompt leptons
- A **fake rate** is derived by comparing *looser* lepton cuts to *tight* lepton cuts used in the main analysis
- **Fake rate** is applied to the application region

Lepton charge mismeasurement

- Also **data-driven**
- **Charge mismeasurement** rate is extracted from the $2\ell ss$ region using $Z/γ^* \rightarrow ee$
- Only applied to the $ee$ region of the analysis
Lepton charge mismeasurement

$2\ell ss$ region using $Z/\gamma^* \rightarrow ee$ used to estimate rate at which the CMS detector incorrectly measures lepton charge

Data-driven

Only applied to the $ee$ region of the analysis
Systematic uncertainties

The systematic uncertainties are: standard and analysis specific

Luminosity – vary simulation by integrated luminosity estimate uncertainty

Jet energy scale (JES) – account for pileup, nonuniform detector response, and any residual differences between the data and simulation

$b$ jet tagging scale factors – account for tagging inefficiencies and charm jet contamination

Cross section theoretical uncertainty – vary cross sections in simulation by uncertainties

PDF shape variations – reweighting the spectra according to the 100 replica sets given by the NNPDF31_NLO_as_0118 PDF parameterization

Renormalization and factorization scale – vary scales by $1/2 \rightarrow 2$

Parton shower – vary ISR by 2 and FSR by $\sqrt{2}$

Matching uncertainty – vary the matching scale between MADGRAPH_aMC@NLO and PYTHIA

Muon and electron ID isolation – vary corrections by their uncertainties

Trigger efficiency – vary corrections by their uncertainties

Pileup – vary the pp inelastic cross section by 5%

Missing parton uncertainty – cover differences between LO EFT w/o extra partons and NLO SM

Uncertainty on the misidentified lepton rate estimate (data-driven) – account for non-prompt contamination

Uncertainty on the charge mismeasurement estimate (data-driven) – account for $Z/\gamma^{*} \rightarrow e^{\pm}e^{\mp}$ becoming $Z/\gamma^{*} \rightarrow e^{\pm}e^{\pm}$
Complete set of 2D contours
Complete set of 2D contours
Changes in event yields over the $2\sigma$ CI

Examining the minimum and maximum yield changes within the $2\sigma$ CI of various WCs