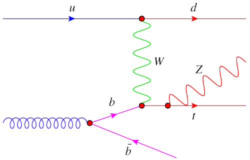
$t\bar{t}Z$, $t\bar{t}W$ and tZq in Run2

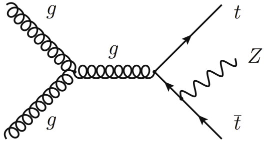
Illia Khvastunov[1], Didar Dobur, Willem Verbeke, Tom Cornelis (UGhent)
Robert Schoefbeck, Daniel Spitzbart (HEPHY)
Andreas Meyer, Joscha Knolle (DESY)
Mirena Paneva (UC Riverside)

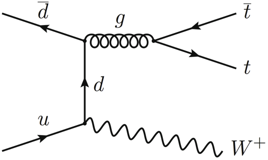
CMS TOP Workshop
November 6, 2018

[1] Also at CEA Saclay

Current status of $t\bar{t}V$ and tZq







TOP-18-008

- Observed (expected) significance is 8.3 (7.2)
- Measured cross section:

$118^{+13}_{-13}(stat.)^{+11}_{-9}(syst.)$fb

TOP-18-009

- Entering the era of differential cross section measurement
- Inclusive cross section:

$1.01^{+0.06}_{-0.05}(stat.)^{+0.07}_{-0.07}(syst.)$pb

TOP-17-005

- Observed (expected) significance is 5.3 (4.5)
- Measured cross section:

$0.77^{+0.12}_{-0.11}(stat.)^{+0.13}_{-0.12}(syst.)$pb

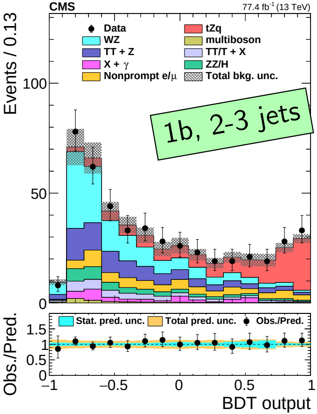
The tZq discovery

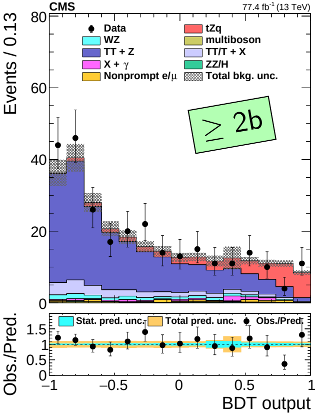
Selection:

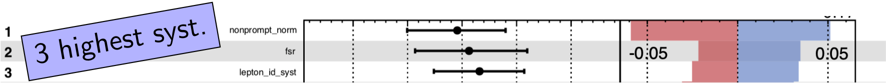
3-lepton channel

3 different BDT in $N_b = 1$, low and high N_j categories, and in $N_b > 1$

Event categorisation in bins of BDT classifier



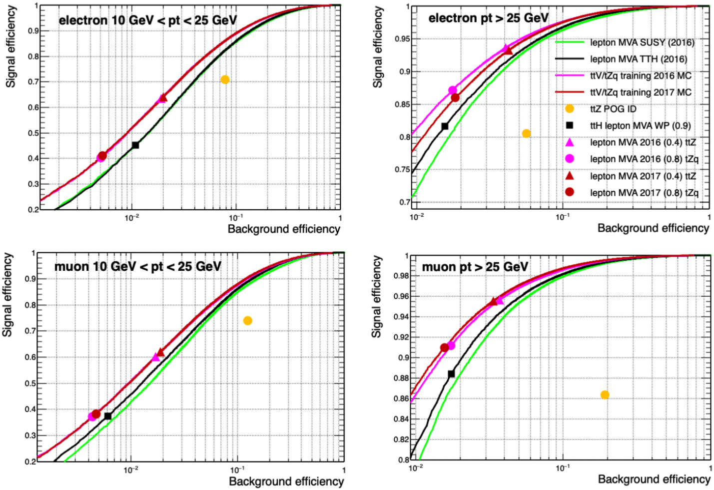




- The crucial ingredient of the discovery is the dedicated Lepton MVA classifier!

Lepton MVA for $t\bar{t}V$ /tZq

- To reduce the nonprompt background in tZq analysis and to suppress the dominant uncertainty in the $t\bar{t}V$ analyses new Lepton MVA discriminator is needed

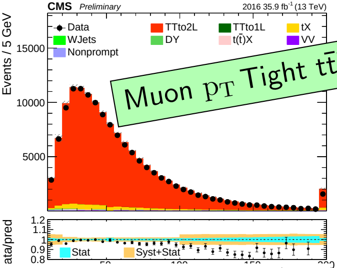


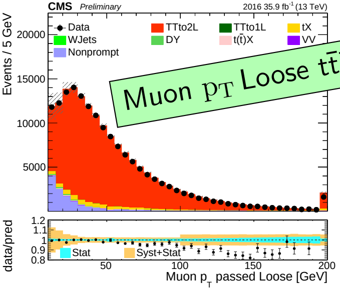
POG recommendation for the Lepton SF uncertainties

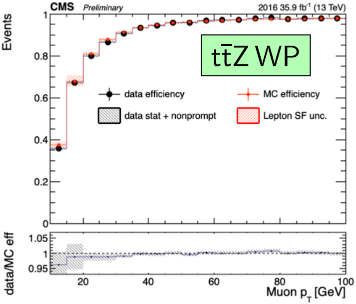
- The TnP code was forked and adjusted for the needs of tZq and $t\bar{t}V$ analysis and the standard procedure was utilised to get the SF
- Numerous detailed studies done on systematic uncertainties, such as:
  - $\rightarrow$ Alternative tag selection and alternative LO DY MC instead of NLO MC, alternative mass shape
  - $\rightarrow$ The most important uncertainty arises from the alternative signal and bkg shape
- The uncertainty ranges from 2-7% for low $\mathrm{p_T}$, 0.5% for intermediate (20-50 GeV) and raises to 1-2% for $\mathrm{p_T} > 100$ GeV
- Muon POG was very interested in the development and performance $\rightarrow$ consider including these WPs in CMSSW for collaboration-wide use

Lepton ID efficiency measurement in $t\bar{t}$ $e\mu$ events

- The goal is to compare efficiencies in $t\bar{t}$ data and MC to verify the extrapolation of the lepton SF measured in Drell-Yan events (study requested by Muon POG)





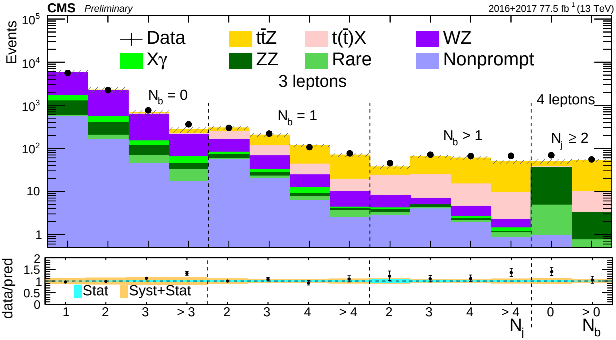


- Good closure in efficiency
- Low p_T bins in the ratio plot close within the systematic uncertainties on the lepton scale factor

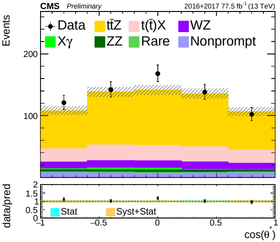
Inclusive and differential $t\bar{t}Z$ x-section measurement

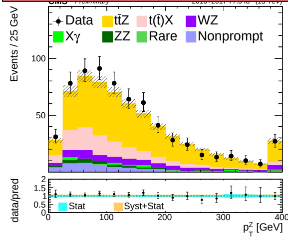
Selection:

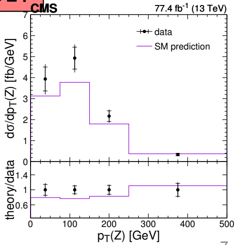
- ⇒ 3- and 4-lepton channel
- ⇒ For inclusive categorisation in N_j and N_b
- ⇒ For differential categorization in p_T^Z and $\cos(\theta^*)$
- ⇒ Unfolding for the first time!



Unblinded for 2016 and 2017







Leading uncertainties in $t\bar{t}Z$

Source	Uncertainty range (%)	Correlated in 2016 and 2017 ?	Impact on the $t\bar{t}Z$ cross section (%)
Trigger	1	$\times$	2
Lepton ID efficiency	4.5-6	✓	4
Parton shower	< 8	✓	3
$t(\bar{t})X$ background	11	✓	3
...			
Total uncertainty			7

- Lepton ID and Trigger uncertainties were addressed in the current analysis
- Next challenge for $t\bar{t}Z$ is the theoretical uncertainties:

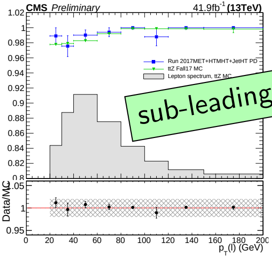
→ new samples with PS weights give us a possibility to estimate them,

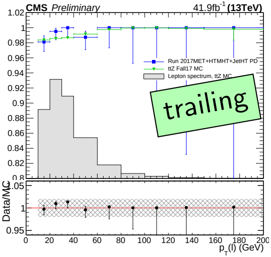
→ for colour reconnection, which is not negligible and affects up to 2%, still have to use $t\bar{t}$ samples

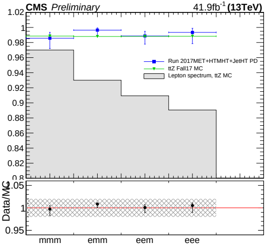
→ Need for dedicated MC samples for theoretical uncertainty estimation

Trigger systematic uncertainty reduction

- The soup of single, di- and trilepton triggers are used
- Efficiency is close to 100%, SF are assigned as a function of leading lepton $\mathrm{p_T}$ (2-3.5%)
- Efficiencies in sub-leading and trailing lepton is checked $\rightarrow$ 1% covers the remaining discrepancy

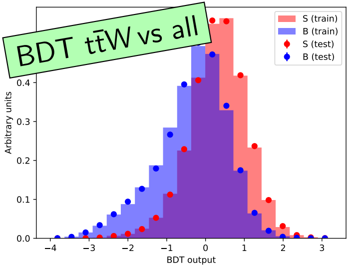


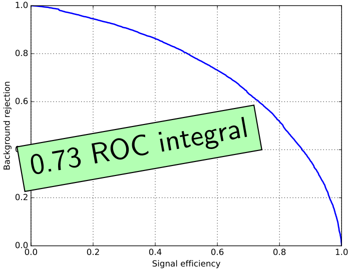




$t\bar{t}W$ challenge

- In previous round of analysis the BDT vs $t\bar{t}$ was trained, achieved discovery level
- Lepton MVA kills the nonprompt background, that was initially done by BDT in TOP-17-005; it enhances as well the yields of $t\bar{t}W$, but in the end no significant improvement was observed from tZq/$t\bar{t}V$ lepton MVA
- New techniques should be used to distinguish between the signal and remaining nonprompt and $t\bar{t}X$ background: new BDT or DNN





- Preliminary results show no significant improvements for DNN wrt to BDT

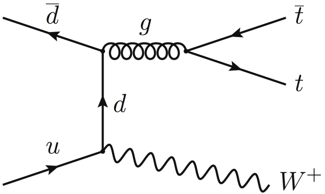
Summary of systematics for $t\bar{t}W$

Source	Uncertainty from each source (%)	Impact on the measured $t\bar{t}W$ cross section (%)
Integrated luminosity	2.5	4
Jet energy scale and resolution	2–5	3
Trigger	2–4	4–5
B tagging	1–5	2–5
PU modeling	1	1
Lepton ID efficiency	2–7	3
Choice in μ_R and μ_F	1	<1
PDF	1	<1
Nonprompt background	30	4
WZ cross section	10–20	<1
ZZ cross section	20	—
Charge misidentification	20	3
Rare SM background	50	2
$t(\bar{t})X$ background	10–15	4
Stat. unc. in nonprompt background	5–50	4
Stat. unc. in rare SM backgrounds	20–100	1
Total systematic uncertainty	—	14

- Lepton SF(decreased to 3%), Trigger (decreased to 2%) and nonprompt background were addressed in current version of $t\bar{t}Z$ analysis
- Effect of the uncertainty from $t\bar{t}X$ should be suppressed by implementing next multivariate discriminator

Next goal for $t\bar{t}W$

- To measure the charge asymmetry (A_c) both in $t\bar{t}W^+$ and $t\bar{t}W^-$
- The A_c is enhanced with respect to $t\bar{t}$ due to selection of $q\bar{q}'$ (LO) and qg (NLO) initial states by the associated W, gg enters only at NNLO (Maltoni et al, arXiv:1406.3262)
- Enhanced A_c^b and A_c^ℓ due to spin correlations and polarization of the $t\bar{t}$ pair ($A_c = \frac{N(|\eta|>|\bar{\eta}|)-N(|\eta|<|\bar{\eta}|)}{N(|\eta|>|\bar{\eta}|)+N(|\eta|<|\bar{\eta}|)}$)



$\Rightarrow A_c^b = 7.54\ ^{+0.19}_{-0.17}\%$
$\Rightarrow A_c^\ell = -13.16\ ^{+1.12}_{-0.81}\%$
$\Rightarrow A_c^t = 2.24\ ^{+0.43}_{-0.32}\%$

- complementary to charge asymmetry measurement in $t\bar{t}$, and differently sensitive to the new physics.
- Back-of-the-envelope calculation shows that the A_c can be probed already in Run 2 with statistical precision of 30% in best case scenario

Conclusions

- **The tZq** is discovered with 2016 and 2017 dataset (TOP-18-008), lepton MVA plays the main role in the discovery; next step is to run EFT interpretations and differential cross section measurement with full Run 2 dataset
- **For $t\bar{t}Z$** 8% uncertainty level has been achieved (TOP-18-009), the analysis became fully systematics limited, new MC samples for better understanding the theoretical uncertainties are needed
- **For $t\bar{t}W$** the discovery level was achieved with 2016 dataset (TOP-17-005), systematic uncertainty can be decreased with current developments from $t\bar{t}Z$ and tZq analyses, next step is to measure charge asymmetry which should be feasible already with Run 2 data
- We started to look into 2018 dataset (have fresh manpower in the group + collaboration with UCLouvain), soon will be presented in one of the meetings

Back-up

Lepton MVA training

- signal: **prompt leptons** in **tZq and $\mathbf{t\bar{t}Z}$** MC
- background: **nonprompt leptons** in combination of **many $\mathbf{t\bar{t}}$ samples**
- **remove leptonic τ decays** from training: τ's are cause a tail in the vertex variables degrading the training
- require muons to mass medium ID, electron to pass dilepton trigger emulation
- $\sim$ 12.3 M prompt electrons, $\sim$ 2.5M nonprompt electrons
- $\sim$ 14.5 M prompt muons, $\sim$ 2.5M nonprompt muons
- 80% of events used for training, 20% for testing

Optimized training parameters and very large training sets lead to very strong discriminator performance

gradient boosted decision tree:

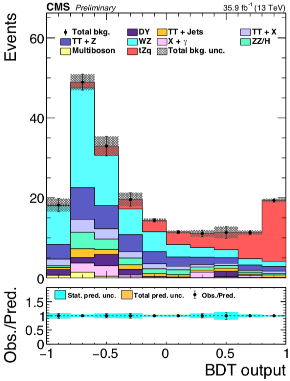
- maximum depth $= 4$
- bagged boost for better generalization

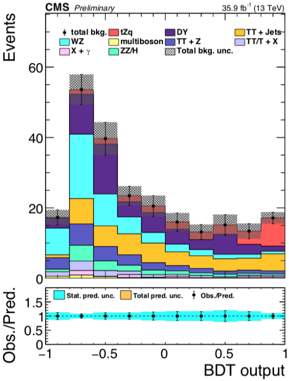
densely-connected feed-forward deep neural network:

- **minor improvement**: $\sim$ 1% higher signal efficiency at similar background rejection compared to BDT
- much more cumbersome to include in CMSSW code $\rightarrow$ keep using BDT
- good **validation of BDT performance**

Replacing lepton MVA with POG selection

with lepton MVA: **TOP-16-020 lepton ID**:





lepton MVA crucial for analysis performance!

- **applied TOP-16-020 lepton selection** in our analysis instead of lepton MVA
- **huge increase in Drell-Yan and TT + jets**
- combined 2016 sensitivity drops **from 6.5** σ **to 2.9** σ, using original BDT training
- after **retraining BDT against the higher fakes** in TOP-16-020 selection $\rightarrow$ **3.9** σ

Muon TnP parameters

Tag selection:

- pass IsoMu24
- combRelIsoPF04dBeta < 0.15
- $p_{\mathrm{T}} > 30$ (in a few specific bins raised with 1 GeV to fix for failed fits)
- Probe multiplicity per tag $== 1$

Probe selection:

- general tracks with $p_{\mathrm{T}} > 10$ GeV

Binning:

- p_{T} bins: (10, 20, 30, 40, 50, 100, 200)
- $|\eta|$ bins: (0., 0.9, 1.2, 2.1, 2.4)

Mass range:

- 70-120 GeV

Shapes used:

- vpvPlusCMS or vpvPlusExpo (depending on which shape worked best for a given p_{T})

Muon systematics

- **alternative tag selection**:
  combRelIsoPF04dBeta < 0.1 and 0.2 instead of 0.15
- **alternative MC sample**:
  using LO DY sample instead of NLO DY
- **alternative mass range**:
  60-130, 70-120 instead of 70-120 GeV
- **alternative background shape**:
  vpvPlusExpo, voigtPlusCMSbeta0p2 or vpvPlusCheb (given what works for a given p_{T} and is different from the nominal one)

Electron TnP parameters

- The **fitting** is based on the old method

  [made my fork already stable and automatized during ICHEP2016 times, so had no need to switch]

  https://github.com/UAEDF-tomc/cmssw/tree/tnp_fitter_94Xtuples/PhysicsTools/TagAndProbe/test

- Calculating scale factors in 2D for p_{T} vs. $|\eta|$
  - p_{T} bins: (10, 20, 30, 40, 50, 100, 200, 500)
  - $|\eta|$ bins: (-2.5, -2.0, -1.566, -1.442, -0.8, 0, 0.8, 1.442, 1.566, 2.0, 2.5)

- Also measurements p_{T} vs. jets, and p_{T} vs. number of primary vertices (for reference, not used as SF)
  - jet bins: (0, 1, 2, 3, 4, 5/6)
  - primary vertices bins: (0, 5, 10, 15, 20, 25, 30, 35, 40, 45, ∞)

- Tag selection: $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.1$, trigger matched (HLT_Ele27_eta2p1_WPTight_Gsf_v* for 2016 or HLT_Ele32_WPTight_Gsf_L1DoubleEG_v for 2017)

- PU weights applied

- All η/p_{T} 2016 fits at http://tomc.web.cern.ch/tomc/tagAndProbe/ttv2018/fits_2016

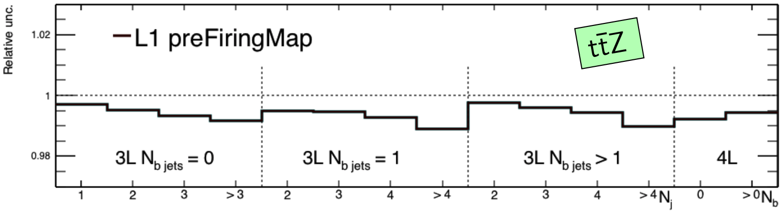
Electron systematics

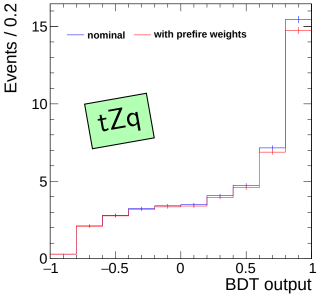
- **nominal**: fit with CMSShape as background and MC templates convoluted with Gaussian as signal
- **alternative tag selection**: $p_{\mathrm{T}} > 35$ GeV, $|\eta| < 2.1$, tag_Ele_trigMVA > 0.92
- **alternative MC sample**: using LO DY sample instead of NLO DY
- **alternative signal shape**: Crystall ball with Gaussian instead of MC templates
- **alternative background shape**: Exponential function instead of CMSShape

Lepton ID efficiency measurement in $t\bar{t}$ $e\mu$ events

- goal is to compare efficiencies in $t\bar{t}$ data and MC to verify the extrapolation of the lepton SF measured in Drell-Yan events
- event selection:
  - → opposite-sign $e\mu$ events, 1 b-jet and ≥ 2 jets
  - → select one good electron tag: pass lepton MVA, $p_T > 40$ GeV
  - → check efficiency of muon passing loose selection to pass lepton MVA cuts
- nonprompt data-driven subtraction was applied:
  - → select same-sign $e\mu$ events, subtract prompt contribution from this,
  - → determine the ratio of number of opposite-sign to same-sign events in MC with at least 1 nonprompt lepton, and multiply same-sign data events by this ratio
  - → subtract respective contamination from numerator and denominator

ECAL L1 pre-firing check in $t\bar{t}Z$ and tZq





⇒ Minor effect on the yields in ttZ, $< 1\%$

⇒ Significant effect in tZq, up to 7% in the most sensitive bin