Search for a single production T’ $\longrightarrow$ top+H in dileptonic SS final state

DESY Meeting, 7 March 2022

Benjamin Blancon[1], Stéphanie Beauceron[1], Candan Dozen-Altuntas[1], Jieun Choi[1 2], Di Wang[3,4]

[1]IP2I, Lyon, France [2]Hanyang University, Seoul, South Korea
[3]Tsinghua University, Beijing, China [4]DESY, Hambourg, Germany

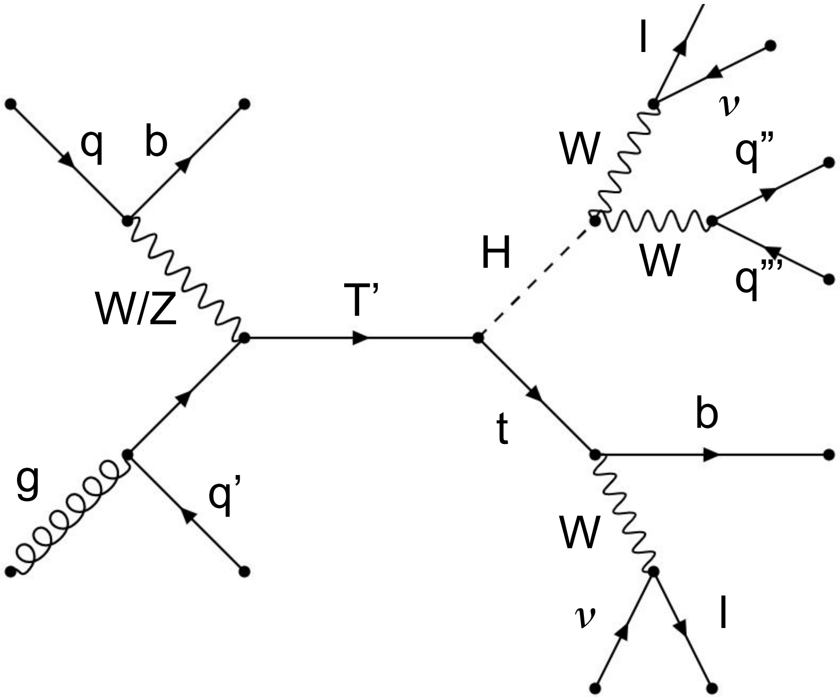
Status of the analysis

New since last presentation:

- Adding the trigger selection.
- Identification of the leptons and the jets.
- Design of the selection criteria.
- Adding the different channels (2μ, 1μ + 1e and 2e).
- Adding all the backgrounds and the mass mass point values only for UL18.
- Design of the control regions for the background estimation method.

Motivation

T' –> Semi-leptonic channel



Vector-Like Quarks (VLQs) are hypothetical particles used in a lot of BSM models to explain the renormalization of the Higgs mass.

The dominant decay mode for single-produced VLQ T' is to third-generation SM quarks: T' ⟶ t+X (t+H in our case).

Previous/ongoing analyses: H ⟶ bb resolved all-hadronic (B2G-18-003/ B2G-19-001), semileptonic (B2G-20-006) and H ⟶ γγ (B2G-21-007) channels.

Datasets (UL18, Luminosity = 150 fb^{-1})

Process	Cross-section (fb)	Number of events	Dataset name
Signal ($M_{T'}$ = 600 GeV)	176.4[1]	400,000	/TprimeBToTH_M-600*
Signal ($M_{T'}$ = 625 GeV)	148.9	400,000	/TprimeBToTH_M-625*
Signal ($M_{T'}$ = 650 GeV)	121.3	276,000	/TprimeBToTH_M-650*
Signal ($M_{T'}$ = 675 GeV)	105.0	400,000	/TprimeBToTH_M-675*
Signal ($M_{T'}$ = 700 GeV)	88.6	400,000	/TprimeBToTH_M-700*
Signal ($M_{T'}$ = 800 GeV)	45.9	397,000	/TprimeBToTH_M-800*
Signal ($M_{T'}$ = 900 GeV)	25.1	400,000	/TprimeBToTH_M-900*
Signal ($M_{T'}$ = 1000 GeV)	14.5	400,000	/TprimeBToTH_M-1000*

[1]: Github for the computation of the theoretical cross-sections (NWA)
* = _LH_TuneCP5_13TeV-madgraph_pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM

Datasets (UL18, Luminosity = 150 fb^{-1})

Process	Cross-section (fb)	Number of events	Dataset name
Signal ($M_{T'}$ = 1100 GeV)	8.67	400,000	/TprimeBToTH_M-1100*
Signal ($M_{T'}$ = 1300 GeV)	3.39	394,000	/TprimeBToTH_M-1300*
Signal ($M_{T'}$ = 1400 GeV)	2.19	400,000	/TprimeBToTH_M-1400*
Signal ($M_{T'}$ = 1500 GeV)	1.45	400,000	/TprimeBToTH_M-1500*
Signal ($M_{T'}$ = 1600 GeV)	0.974	367,000	/TprimeBToTH_M-1600*
Signal ($M_{T'}$ = 1700 GeV)	0.663	400,000	/TprimeBToTH_M-1700*
Signal ($M_{T'}$ = 1800 GeV)	0.458	400,000	/TprimeBToTH_M-1800*

No 1200 GeV mass point dataset for UL18.

* = _LH_TuneCP5_13TeV-madgraph_pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM

Datasets (UL18, Luminosity = 150 fb^{-1})

Process	Cross-section (fb)	Number of events	Dataset name
ttW	610.5	27,686,862	/ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8/*-v2/NANOAODSIM
ttZ	770	32,793,815	/ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8/*-v2/NANOAODSIM
ttH	271	7,328,993	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM
ttbar semileptonic	364,351	476,408,000	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM
ttbar dileptonic	87,315	145,020,000	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM
Single top (tW)	19,467	22,220,050	/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM /ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM
WZ	4,429.7	9,821,283	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v2/NANOAODSIM
ZZ	1,256	98,488,000	/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/*-v2/NANOAODSIM

* = RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1

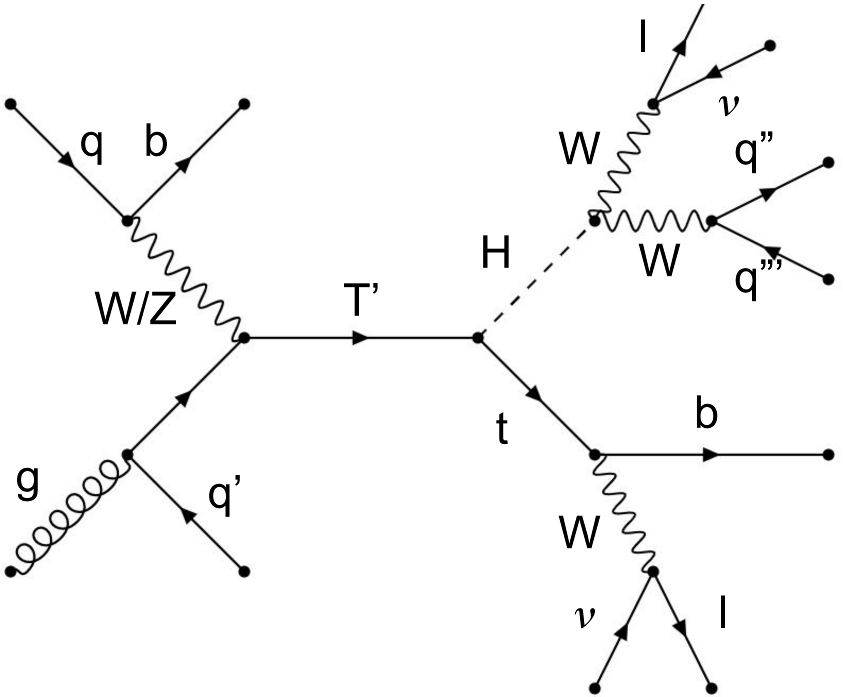
Datasets (UL18, Luminosity = 150 fb^{-1})

Process	Cross-section (fb)	Number of events	Dataset name
WWW	208.6	240,000	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM
WH	31.3	19,916,695	/HWminusJ_HToWWTo2L2Nu_WTo2L_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM /HWplusJ_HToWWTo2L2Nu_WTo2L_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM
ZH	185.8	9,899,256	/HZJ_HToWW_M-125_TuneCP5_13TeV-powheg-jhugen727-pythia8/*-v2/NANOAODSIM
WW	12,178	9,994,000	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM
WW Deep Scattering	170.3	500,000	/WWTo2L2Nu_TuneCP5_DoubleScattering_13TeV-pythia8/*-v2/NANOAODSIM
Drell-Yan	18,610,000 (10-50 GeV) 6,020,850 (50 GeV)	94,452,816 96,233,328	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM /DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM
W+Jets	61,334,900	81,051,269	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM

* = RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1

Analysis strategy

T' –> Semi-leptonic channel



Analysis focused in the dileptonic (2µ, 1µ + 1e and 2e) SS channel. The final state has three additional jets with one b-jet.

Looking for a bump in the St (= Sum of Pt of the two leptons + all the jets) spectrum.

Background estimation extracted from the data.

Determination of background shape thanks to combined fits of Signal Region and Control Regions and extraction of signal shape.

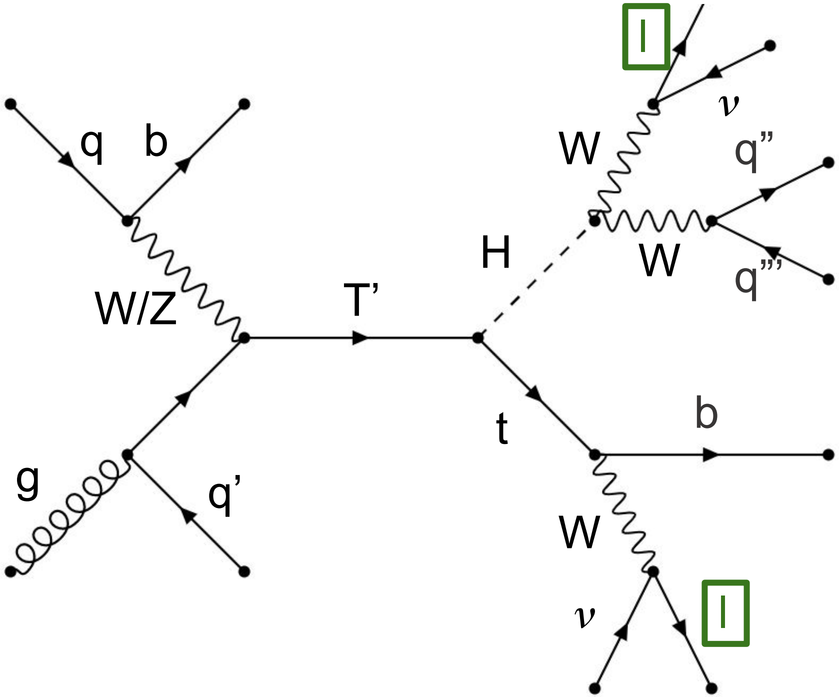
Trigger selection

List of the trigger HLT paths for the different channels:

- HLT_IsoMu24.
- HLT_Ele32_WPTight_Gsf.
- HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8.
- HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ.
- HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL.
- HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL.

Object identification

T' –> Semi-leptonic channel



Lepton identification:

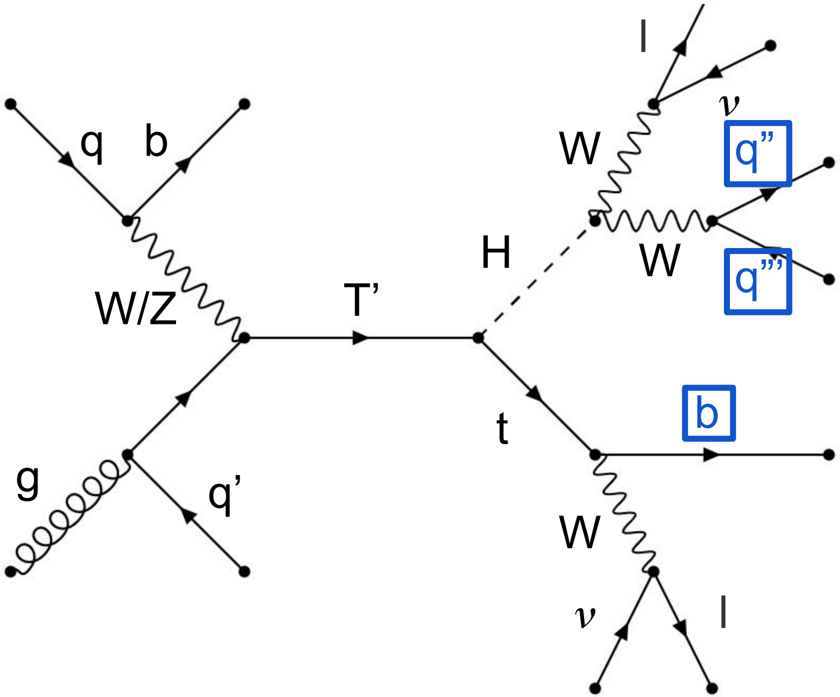
- electron: Pt > 25 GeV, |η| < 2.5, tight ID, isolation[1] < 0.05, 3D Impact parameter significance sip3D < 2 cm and tight charge = 2 to reject charge flip.
- muon : Pt > 20 GeV, |η| < 2.4, tight ID, isolation[1] < 0.05, sip3D < 3 cm and tight charge = 2.

Veto for an additional lepton.

[1]: mini-isolation (Section II.A.)

Object identification

T' –> Semi-leptonic channel



Jet identification:

- jet: Pt > 30 GeV, |η| < 4.5, tight ID.
- b-jet: Pt > 30 GeV, |η| < 2.5, tight ID and medium DeepJet WP.

Cleaning of the jets in a cone of ΔR < 0.4 with loose leptons:

- loose electron: Pt > 25 GeV, |η| < 2.5, loose ID, isolation[1] < 0.10, sip3D < 2 cm and tight charge = 2.
- loose muon: Pt > 20 GeV, |η| < 2.4, tight ID, isolation[1] < 0.40, sip3D < 3 cm and tight charge = 2.

[1]: mini-isolation (Section II.A.)

Selection criteria

- Cut 0: exactly 2 SS leptons with the sum of Pt of two leptons > 160 GeV.
- Cut 1: ≥ 3 jets and ≥ 1 b-jet with Pt > 50 GeV.
- Cut 2: Min(3 jets invariant mass - top invariant mass) > 34 GeV (i.e. 3 jets invariant mass ∉ top invariant mass ± 2σ[1]).
- Cut 3: ΔR(leptons) > 1.8.

	Signal (600 GeV)	Signal (700 GeV)	Signal (800 GeV)	Signal (900 GeV)	Signal (1 TeV)	Signal (1.1 TeV)	Sum of backgrounds
Cut 0	34.1	19.6	12.0	7.7	4.5	2.8	4476.0
Cut 1	18.6	11.5	7.2	4.7	2.7	1.7	631.0
Cut 2	14.6	8.9	6.2	4.1	2.4	1.5	351.3
Cut 3	13.1	8.6	5.9	4.0	2.3	1.5	255.2

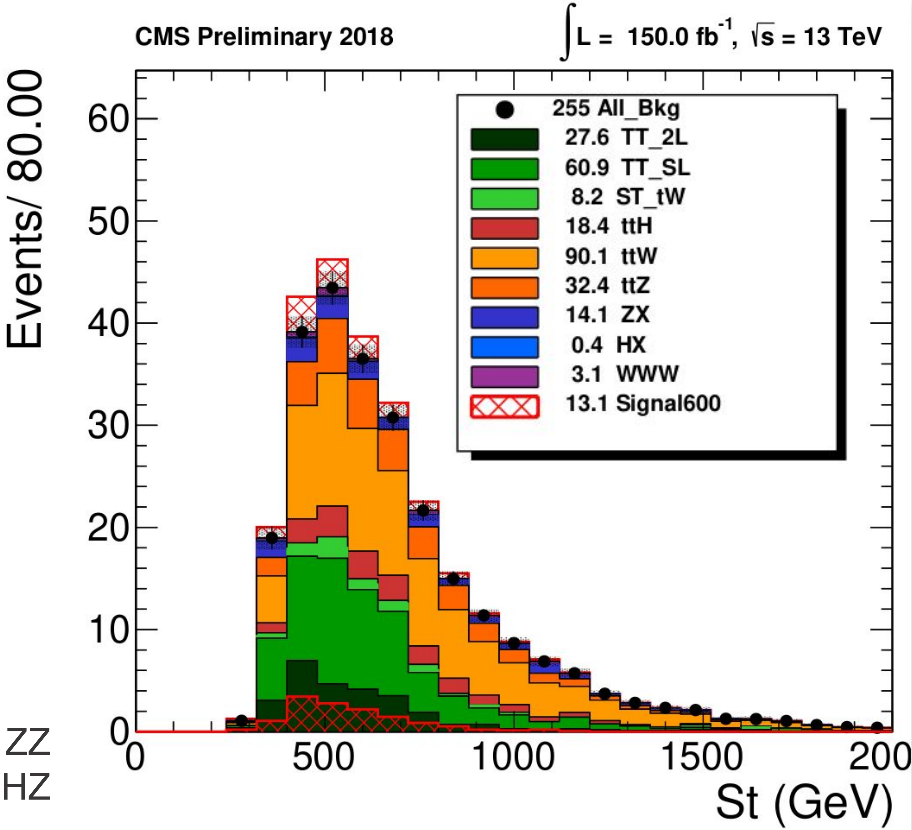
All the channels are considered. The signal is negligible above 1 TeV (see backup slides for all the mass point values).

[1]: B2G-19-001, section 3.2.2

St distribution (T' mass = 600 GeV)

Reminder: St = Sum of Pt of the two leptons + jets.

ttX (ttH + ttW + ttZ) (55%), ttbar semileptonic (TT_SL) (24%) and ttbar dileptonic (TT_2L) (11%) are the main backgrounds. The signal is visible but the peak is at the same place as the background.



ZX = ZW + ZZ
HX = HW + HZ

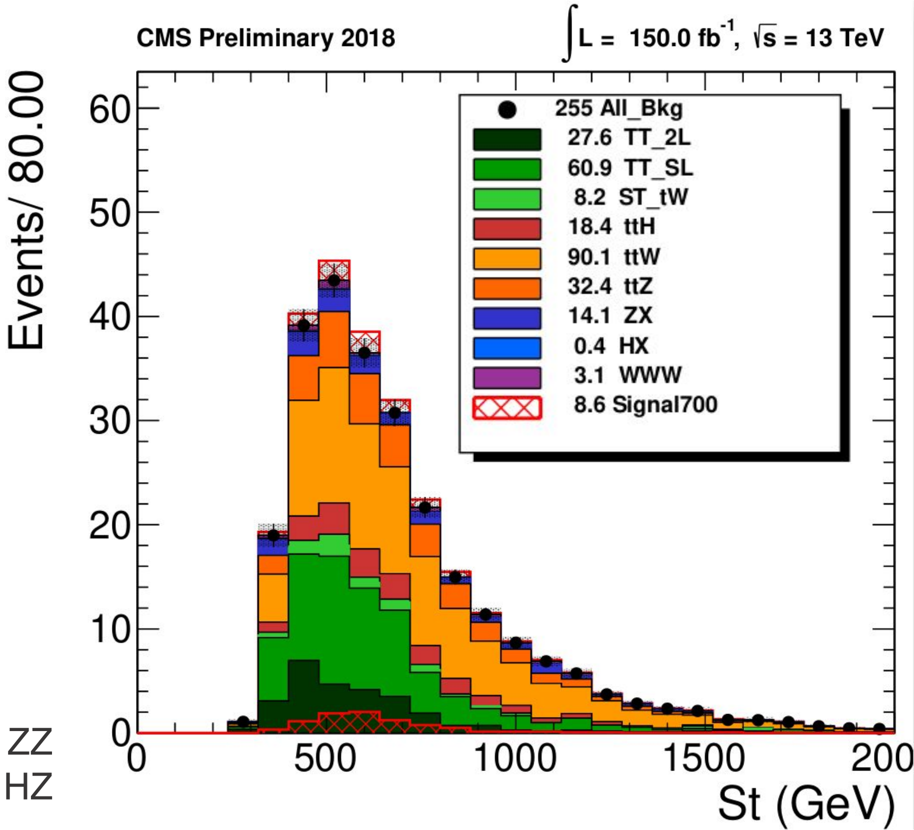
St distribution (T’ mass = 700 GeV)

Reminder: St = Sum of Pt of the two leptons + jets.

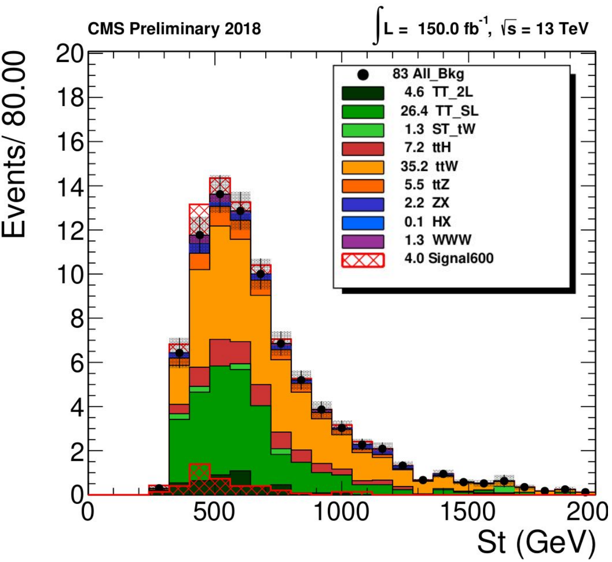
ttX (55%), TT_SL (24%) and TT_2L (11%) are the main backgrounds. The signal peak is shifted but is barely visible (getting worse as the T’ mass increases).

Using the GetQuantiles method? Study is ongoing.

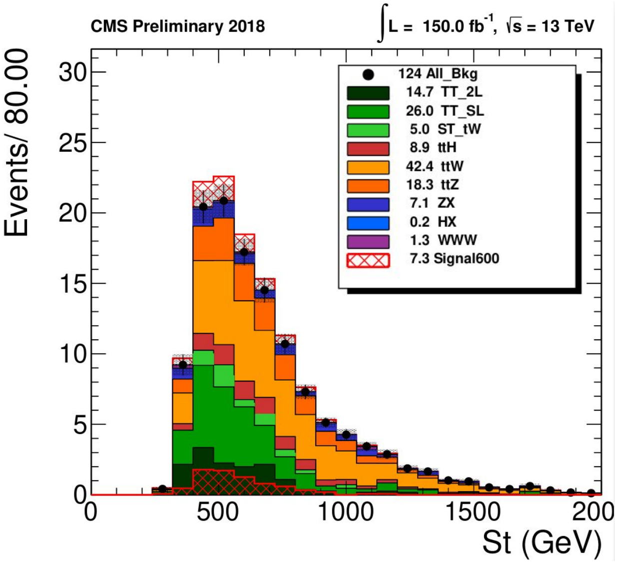
ZX = ZW + ZZ
HX = HW + HZ



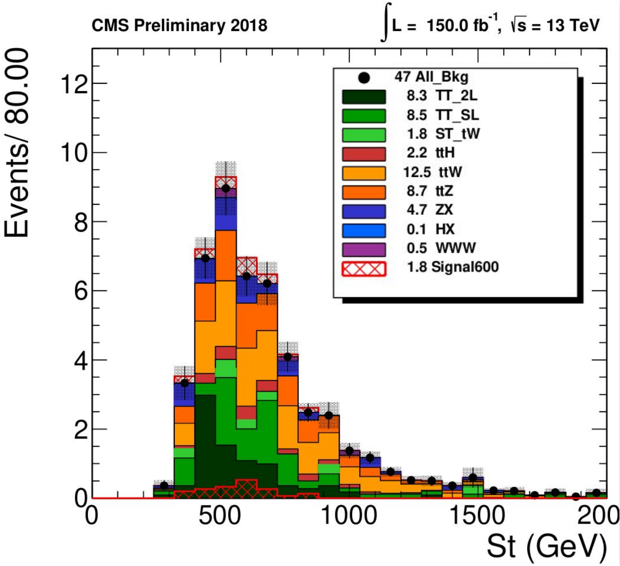
St distribution (T’ mass = 600 GeV, separated channels)



2μ channel



1μ+1e channel



2e channel

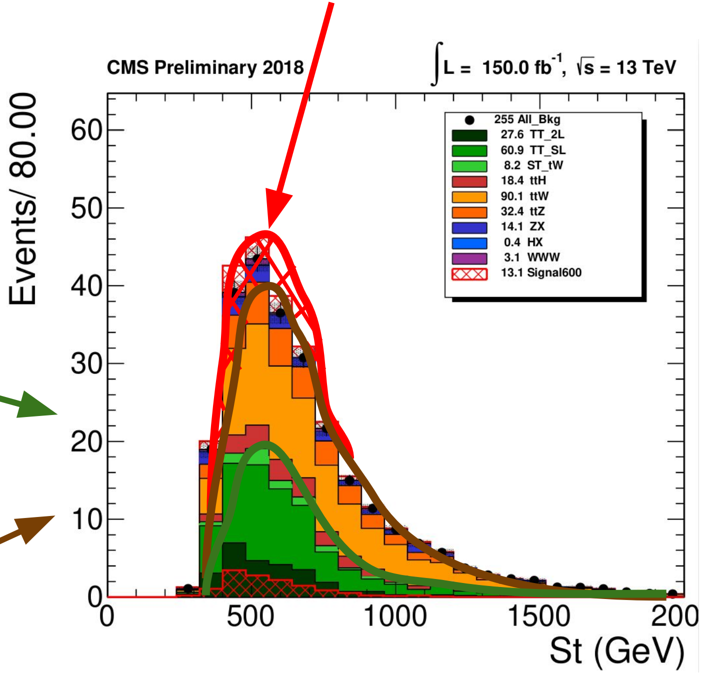
The signal is visible in the muonic channels but the signal peak still clashes with the background. The signal is not visible in the dielectronic channel because of the complexity to correctly identify the electrons.

Benjamin Blancon
DESY Meeting
03.07.23

Control regions

Goal ⟶ Define background shape thanks to Control regions (CRs) and apply it to the Signal Region (SR) to extract signal shape.

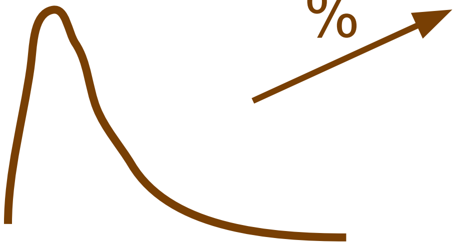
Signal shape in SR



CR 1: first background shape

%

CR 2: second background shape



%

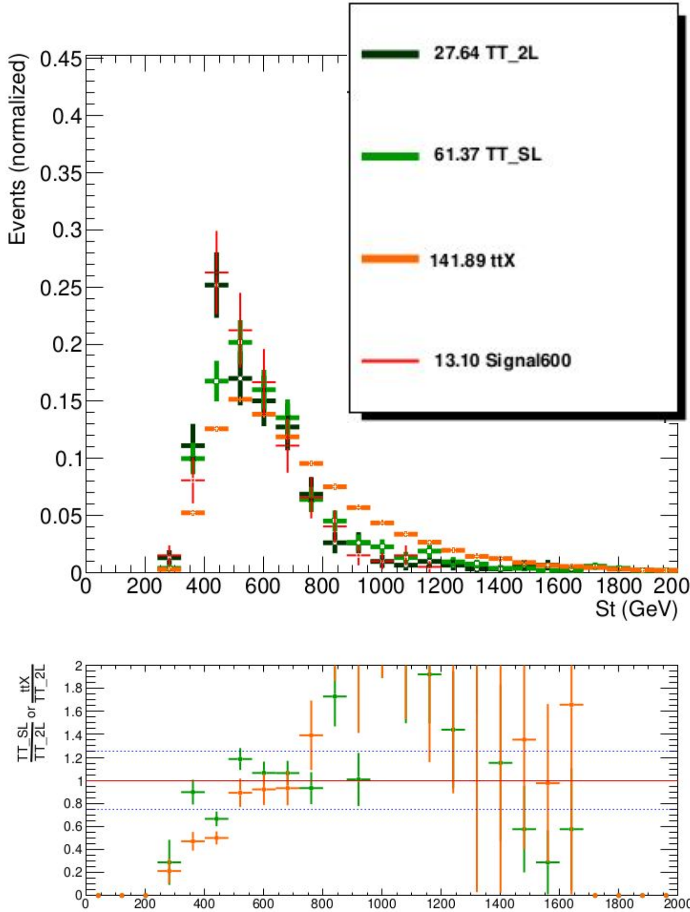
…

Control regions

The ratio is computed wrt TT_2L. ttX, TT_SL and TT_2L have three different shapes for St in the SR (we consider the T' mass = 600 GeV for the next slides).

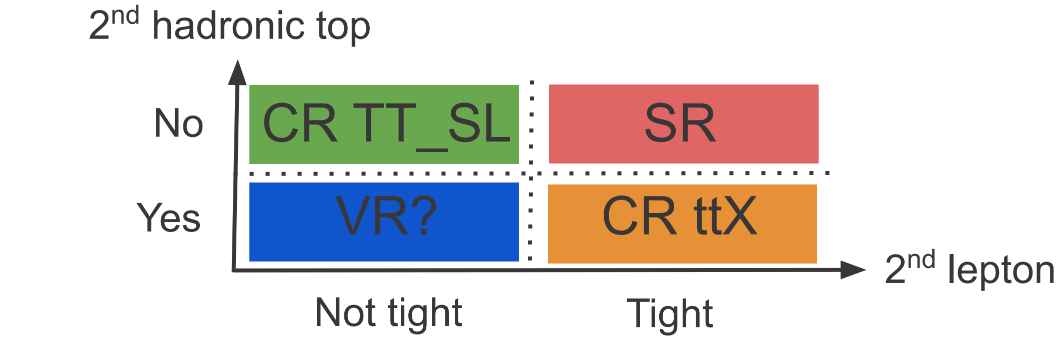
⟶ Three CRs by inverting one cut:

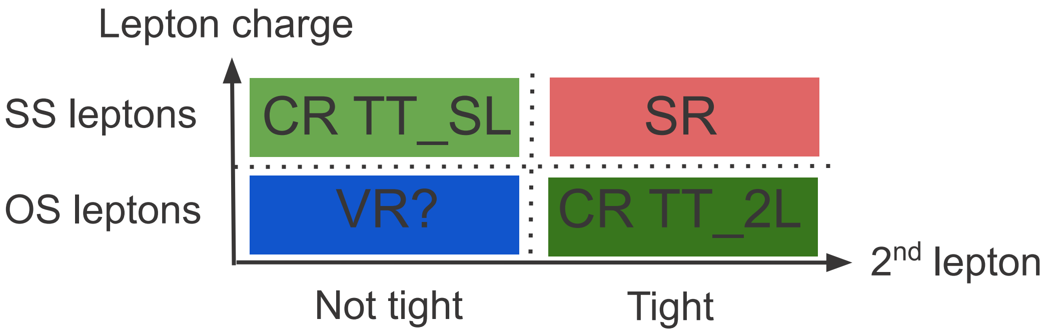
- ttX: at least one hadronic top decay.
- TT_SL: one loose lepton that fails the tight criteria and one tight lepton.
- TT_2L: two leptons opposite sign.

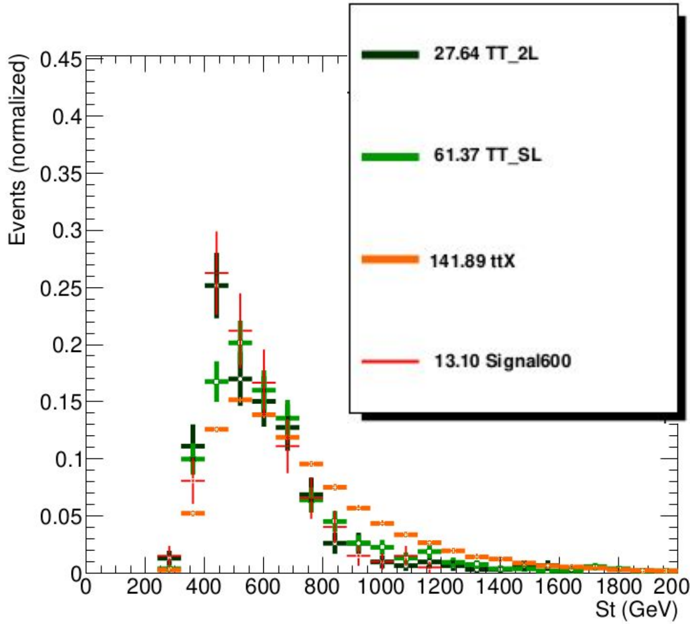


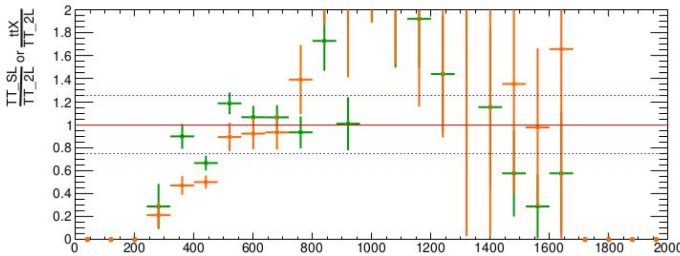
Benjamin Blancon
B2G General Meeting
09.27.22

Control regions



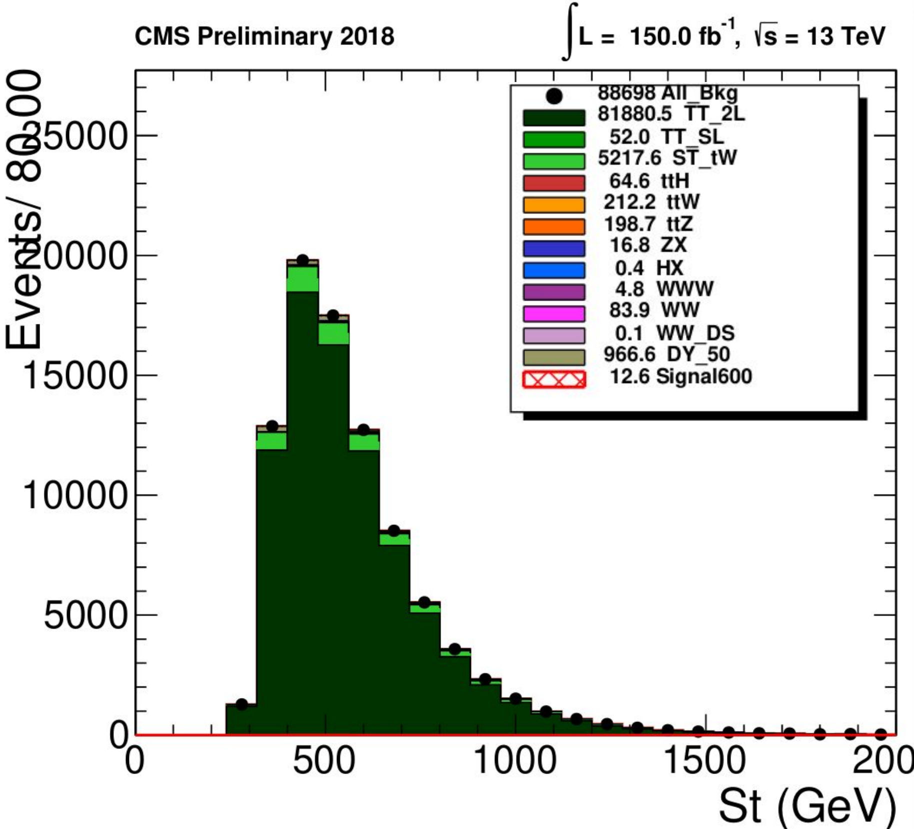






Control regions (ttbar dileptonic)

TT_2L is the largest background in its CR (92.3%). ttX, TT_SL and the signal contamination are negligible.
⟶ OK!

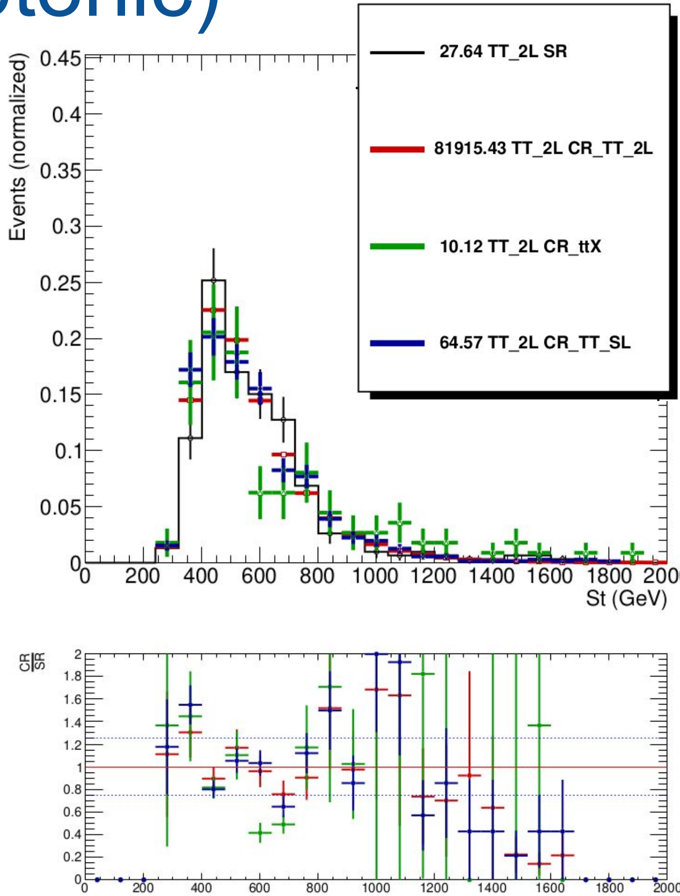


Control regions Validation (ttbar dileptonic)

The ratio plot is computed wrt SR. TT_2L is negligible in the CR ttX (see next slides). The shapes seem different in the SR, the CRs TT_2L and TT_SL (the systematics have to be added).

⟶ Lack of statistics?

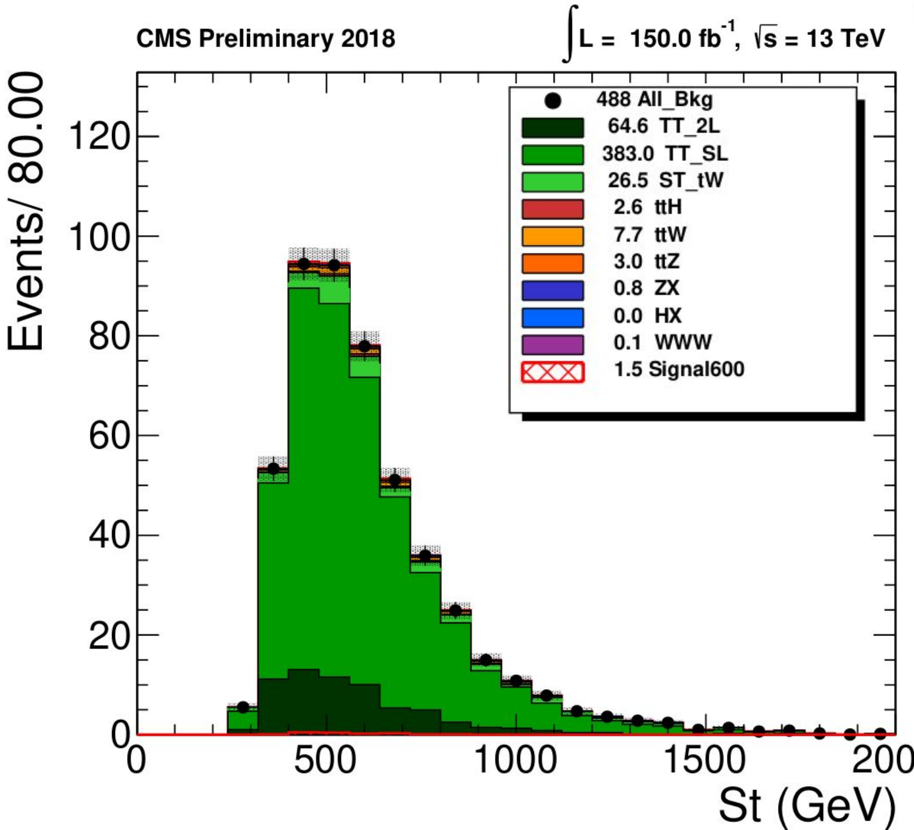
⟶ Valid method for TT_2L(?)



Control regions (ttbar semileptonic)

TT_SL is the largest background in its CR (78.4%). ttX and the signal contamination are negligible. TT_2L is not negligible but it has the same shape in the CRs TT_2L and TT_SL.

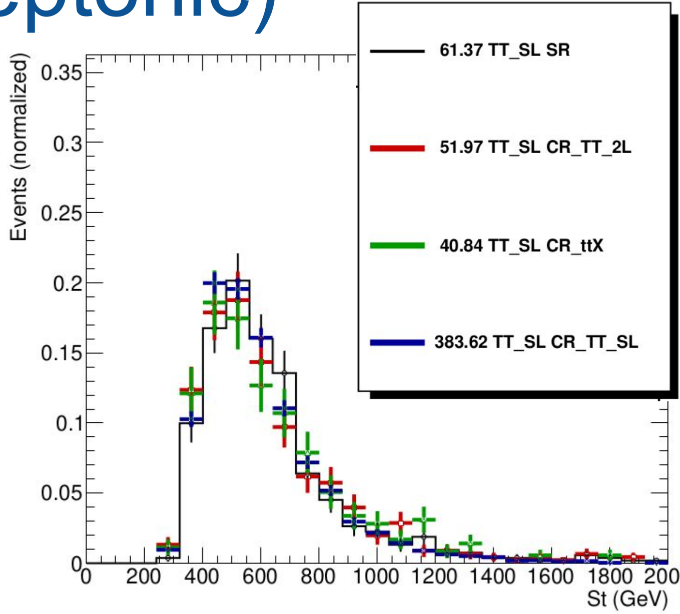
⟶ OK!

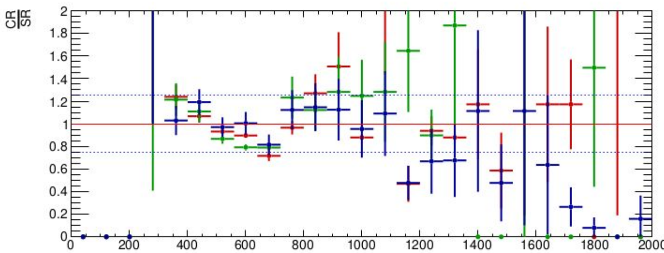


Control regions
Validation (ttbar semileptonic)

The ratio plot is computed wrt SR. TT_2L has the same shape in the SR and its CR (the systematics have to be added). The shape is more tricky for the CR ttX. ⟶ Study is ongoing.

⟶ Valid method for TT_SL(?)

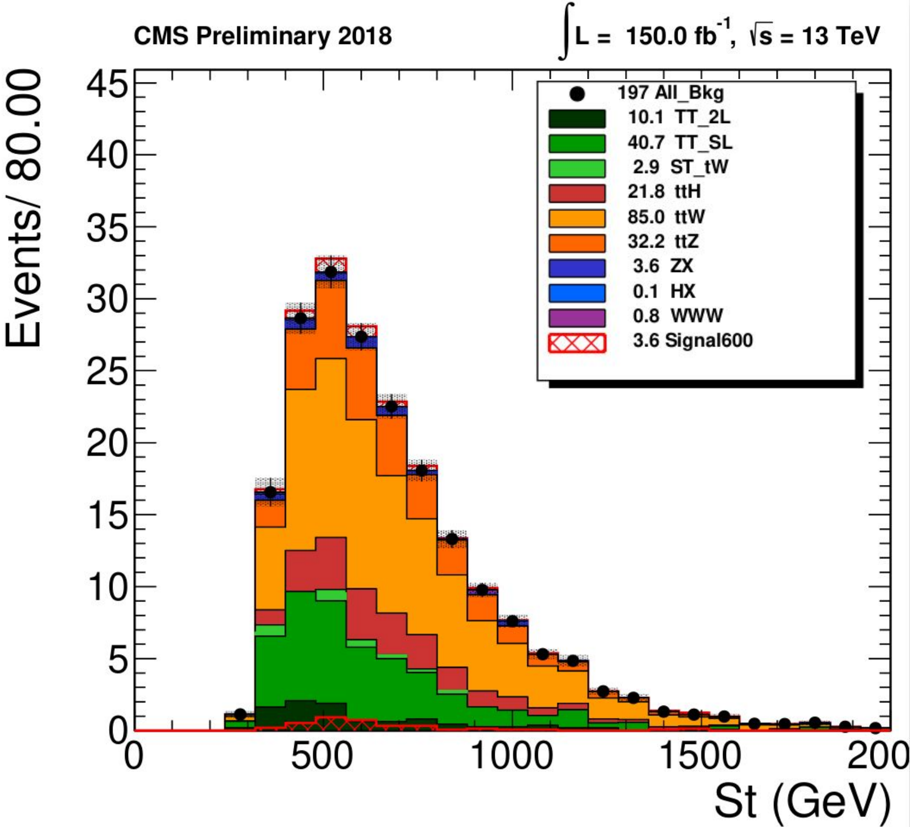




Control regions (ttX)

ttX is the largest background in its CR (70.8%). TT_SL is not negligible (20.7%) but we assume its shape is the same in the CRs TT_SL and ttX.

⟶ OK!

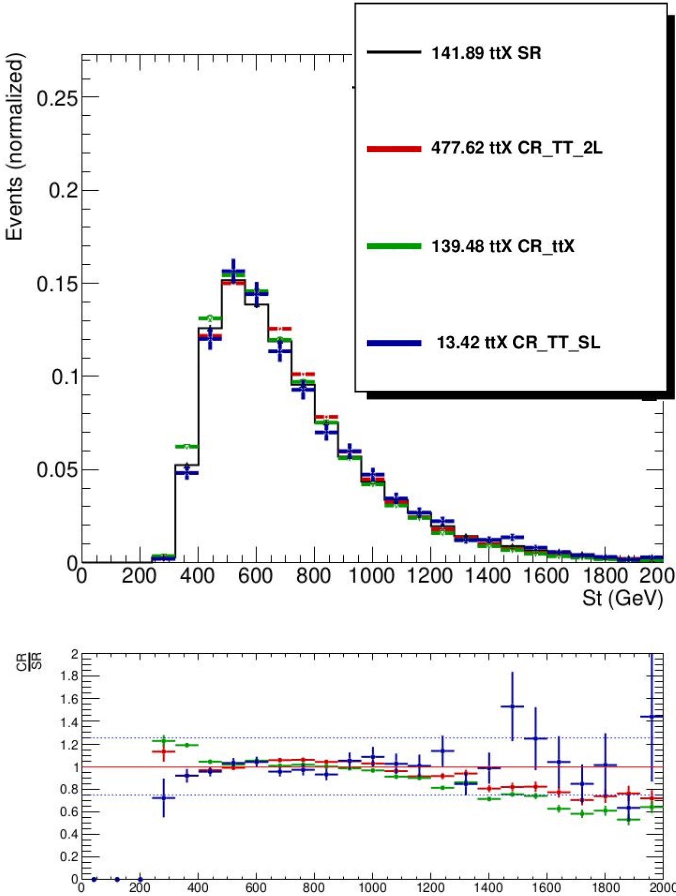


Benjamin Blancon
DESY Meeting
03.07.23

Control regions Validation (ttX)

The ratio plot is computed wrt SR. ttX has the same shape in the SR and its CR (the systematics have to be added).

→ Valid method for ttX!



Benjamin Blancon
DESY Meeting
03.07.23

Results with Combine


```
Combination of SR=datacard_SR.txt  CR_TT_2L=datacard_CR_TT_2L.txt  CR_ttX=datacard_CR_ttX.txt  CR_TT_SL=datacard_CR_TT_SL.txt
imax 4 number of bins
jmax 3 number of processes minus 1
kmax 0 number of nuisance parameters
----------------------------------------------------------------------------------------------------------------------------------
shapes TT_2L     CR_TT_2L  workspace_data_Signal600.root w:TT_2L_CR_TT_2L_histo
shapes TT_SL     CR_TT_2L  workspace_data_Signal600.root w:TT_SL_CR_TT_2L_histo
shapes data_obs  CR_TT_2L  workspace_data_Signal600.root w:Data_CR_TT_2L_histo
shapes signal    CR_TT_2L  workspace_data_Signal600.root w:Signal_CR_TT_2L_histo
shapes ttX       CR_TT_2L  workspace_data_Signal600.root w:ttX_CR_TT_2L_histo
shapes TT_2L     CR_TT_SL  workspace_data_Signal600.root w:TT_2L_CR_TT_SL_histo
shapes TT_SL     CR_TT_SL  workspace_data_Signal600.root w:TT_SL_CR_TT_SL_histo
shapes data_obs  CR_TT_SL  workspace_data_Signal600.root w:Data_CR_TT_SL_histo
shapes signal    CR_TT_SL  workspace_data_Signal600.root w:Signal_CR_TT_SL_histo
shapes ttX       CR_TT_SL  workspace_data_Signal600.root w:ttX_CR_TT_SL_histo
shapes TT_2L     CR_ttX    workspace_data_Signal600.root w:TT_2L_CR_ttX_histo
shapes TT_SL     CR_ttX    workspace_data_Signal600.root w:TT_SL_CR_ttX_histo
shapes data_obs  CR_ttX    workspace_data_Signal600.root w:Data_CR_ttX_histo
shapes signal    CR_ttX    workspace_data_Signal600.root w:Signal_CR_ttX_histo
shapes ttX       CR_ttX    workspace_data_Signal600.root w:ttX_CR_ttX_histo
shapes TT_2L     SR        workspace_data_Signal600.root w:TT_2L_histo
shapes TT_SL     SR        workspace_data_Signal600.root w:TT_SL_histo
shapes data_obs  SR        workspace_data_Signal600.root w:Data_histo
shapes signal    SR        workspace_data_Signal600.root w:Signal_histo
shapes ttX       SR        workspace_data_Signal600.root w:ttX_histo
----------------------------------------------------------------------------------------------------------------------------------
bin          SR        CR_TT_2L  CR_ttX    CR_TT_SL
observation  -1        -1        -1        -1
----------------------------------------------------------------------------------------------------------------------------------
bin          SR        SR        SR        SR        CR_TT_2L  CR_TT_2L  CR_TT_2L  CR_TT_2L  CR_ttX    CR_ttX    CR_ttX    CR_ttX    CR_TT_SL  CR_TT_SL  CR_TT_SL  CR_TT_SL
process      signal    TT_2L     ttX       TT_SL     signal    TT_2L     ttX       TT_SL     signal    TT_2L     ttX       TT_SL     signal    TT_2L     ttX       TT_SL
process      0         1         2         3         0         1         2         3         0         1         2         3         0         1         2         3
rate         -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1        -1
----------------------------------------------------------------------------------------------------------------------------------

scale_TT_2L rateParam CR_TT_2L TT_2L 1 [0,2]
scale_TT_2L rateParam CR_TT_SL TT_2L 1 [0,2]
scale_TT_SL rateParam CR_TT_SL TT_SL 1 [0,2]
# scale_TT_2L rateParam CR_ttX TT_2L 1 [0,2]
scale_TT_SL rateParam CR_ttX TT_SL 1 [0,2]
scale_ttX rateParam CR_ttX ttX 1 [0,2]
scale_TT_2L rateParam SR TT_2L 1 [0,2]
scale_TT_SL rateParam SR TT_SL 1 [0,2]
scale_ttX rateParam SR ttX 1 [0,2]
```

Three backgrounds considered: TT_2L, TT_SL and ttX.

Three parameters to have the same background shape in the regions:

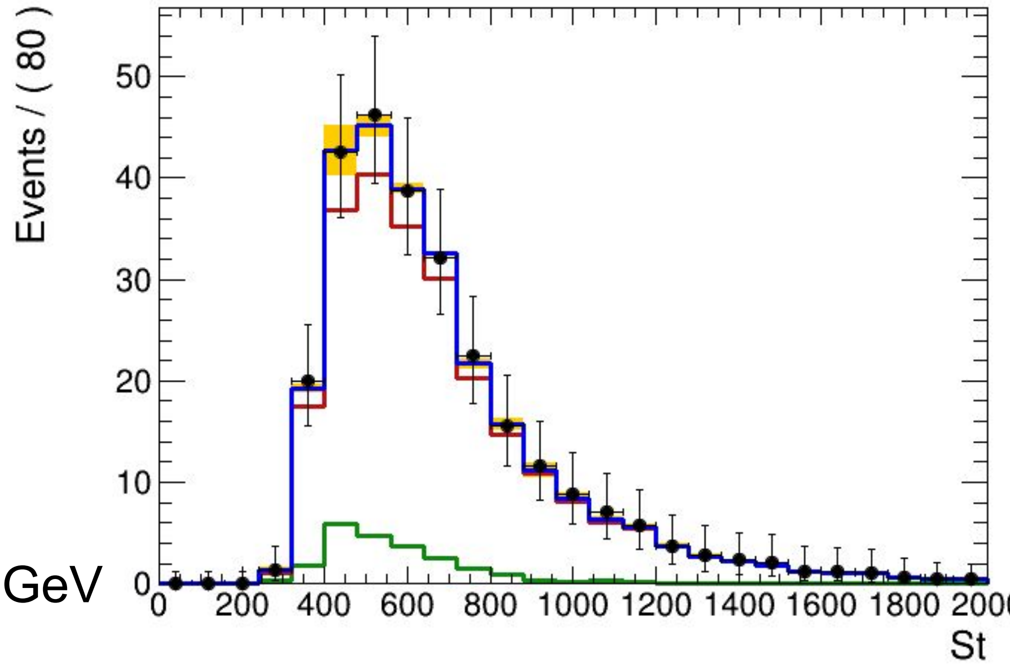
- Scale_TT_2L in CR TT_2L, CR TT_SL and SR.
- Scale_TT_SL in CR TT_SL, CR ttX and SR.
- Scale_ttX in CR ttX and SR.

Results with Combine

POI r and Significance σ (see backup slides for all the mass point values)

T' Mass (GeV)	POI r	Significance σ
600	1.72 ± 1.55	1.132
700	2.30 ± 2.59	0.872
800	2.49 ± 3.22	0.735
900	4.12 ± 4.91	0.764
1000	5.29 ± 7.13	0.648
1100	7.5 ± 13.6	0.720

600 GeV



I'm still not sure how to interpret these values…?

To do list

- Finish the design for the CRs.
- Check for the GetQuantiles method (see Candan's talk).
- Implement the weights.
- First look at the data (2018 UL).
- Add the full Run 2 statistics.
- Start writing the Analysis Note.

Summary

We focused on T' $\longrightarrow$ top+H dileptonic SS final state analysis.

Analysis strategy designed: multiple studies (lepton identification, cut selection).

Defined CRs for the background estimation method.

We expect a data over MC comparison soon!

Thank you!

Back-up

Selection criteria

	Signal (600 GeV)	Signal (625 GeV)	Signal (650 GeV)*	Signal (675 GeV)	Signal (700 GeV)	Signal (800 GeV)	Signal (900 GeV)	Signal (1 TeV)	Sum of backgrounds
Cut 0	34.1	27.9	24.0	25.2	19.6	12.0	7.7	4.5	4476.0
Cut 1	18.6	15.5	11.9	13.5	11.5	7.2	4.7	2.7	631.0
Cut 2	14.6	12.9	9.4	11.1	8.9	6.2	4.1	2.4	351.3
Cut 3	13.1	11.9	8.8	10.5	8.6	5.9	4.0	2.3	255.2

	Signal (1.1 TeV)	Signal (1.3 TeV)	Signal (1.4 TeV)	Signal (1.5 TeV)	Signal (1.6 TeV)	Signal (1.7 TeV)	Signal (1.8 TeV)	Sum of backgrounds
Cut 0	2.8	1.1	0.7	0.5	0.3	0.2	0.1	4476.0
Cut 1	1.7	0.7	0.4	0.3	0.2	0.1	0.1	631.0
Cut 2	1.5	0.6	0.4	0.2	0.2	0.1	0.1	351.3
Cut 3	1.5	0.6	0.4	0.2	0.2	0.1	0.1	255.2

* Lack of statistics for 650 GeV.

Results with Combine

	Signal (600 GeV)	Signal (625 GeV)	Signal (650 GeV)	Signal (675 GeV)	Signal (700 GeV)	Signal (800 GeV)	Signal (900 GeV)	Signal (1 TeV)
ROI r	1.72 ± 1.55	2.54 ± 2.92	2.31 ± 2.50	2.13 ± 2.10	2.30 ± 2.59	2.49 ± 3.22	4.12 ± 4.91	5.29 ± 7.13
Significance σ	1.132	0.844	0.902	0.998	0.872	0.735	0.764	0.648

	Signal (1.1 TeV)	Signal (1.3 TeV)	Signal (1.4 TeV)	Signal (1.5 TeV)	Signal (1.6 TeV)	Signal (1.7 TeV)	Signal (1.8 TeV)
ROI r	7.5 ± 13.6	18.5 ± 11.2	20.0 ± 15.8	20.0 ± 19.7	20.0 ± 19.5	20.0 ± 19.2	20.0 ± 16.0
Significance σ	0.720	0.757	0.737	0.821	0.858	0.979	0.869