Search for charged Higgs boson with the $H^{\pm} \to tb$ decay in fully hadronic final state

B2G Resonances meeting

A. Attikis[1], K. Christoforou[1], M. Kolosova[1], **S. Konstantinou**[1], S. Lehti[2], C.Leonidou[1], L.Paizanos[1], F. Ptochos[1], H. Saka[1], A. Stepennov[1]

[1]UCY [2]HIP

Friday 25th November, 2022

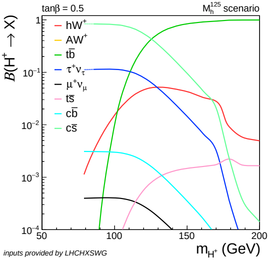
$H^{\pm}$ predicted by many BSM theories that extend their Higgs sector

- two-Higgs-doublet models (2HDMs) predict 5 physical states:
  - two $\mathcal{CP}$-even h^0 and H^0, $\mathcal{CP}$-odd A^0, two $H^{\pm}$

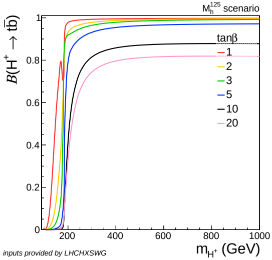
Three mass categories are commonly defined in $H^{\pm}$ searches:

- Light $m_{H^{\pm}} < m_t - m_b$, intermediate $m_{H^{\pm}} \sim m_t$, heavy $m_{H^{\pm}} > m_t + m_b$

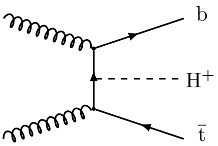
M_h^{125}, $\tan\beta = 0.5$



M_h^{125}, $H^{\pm} \to tb$



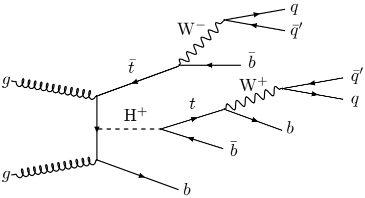
heavy $m_{H^{\pm}}$ (4FS)



BRs of $H^{\pm} \to tb$ dominanates at high $m_{H^{\pm}}$, for wide range of $\tan\beta$

Fully-hadronic final state of associated production characterised by:

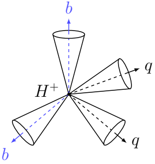
- High jet & b jet multiplicities
- ✓ Large branching ratio $\mathcal{B} \simeq 46\%$
- ✓ Invariant mass reconstruction of $\mathrm{H}^{\pm}$
- ✗ QCD multijet & $\mathrm{t\bar{t}}$ background
- ✗ Combinatorial (self-)background



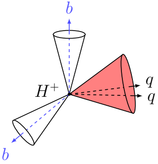
Various $\mathrm{m}_{\mathrm{H}^{\pm}}$ reconstruction techniques available due to signal process kinematics:

- **Resolved t**: At moderate $m_{\mathrm{H}^{\pm}}$ & $p_{\mathrm{T},\mathrm{H}^{\pm}}$ the decay products of $\mathrm{H}^{\pm}$ are well separated
- **Boosted W/t**: As $\mathrm{m}_{\mathrm{H}^{\pm}}$ increases the $\mathrm{H}^{\pm}$ decay products become boosted

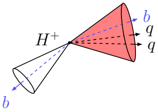
resolved t



boosted W



boosted t



Previous results

- **Resolved t**, **Boosted W/t** studied separately by dedicated analyses
- 2016 ReReco data
- CADI HIG-18-015

This work

- 3 main categories
  - **2 resolved t**
  - **1 resolved, 1 boosted t** NEW
  - **2 boosted t**
- Last report (B2G-RES): 12 Nov 2022

- This analysis targets full Run II data
- This talk presents a study using 2018 data

Datasets	**Luminosity** (pb^{-1})
JetHT_Run2018A_UL2018_MiniAODv2_v1_315257_316995	14026.95
JetHT_Run2018B_UL2018_MiniAODv2_v1_317080_319310	7060.79
JetHT_Run2018C_UL2018_MiniAODv2_v1_319337_320065	6894.78
JetHT_Run2018D_UL2018_MiniAODv2_v2_320413_325172	31834.89
Total:	59817.41

MC simulated samples include:

- Signal: $m_{\mathrm{H}^{\pm}} = 200 - 3000$ GeV (17 points)
- QCD (H_{T} binned)
- Top (Single top, $\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}}$ +X)
- V+jets, diboson, triboson

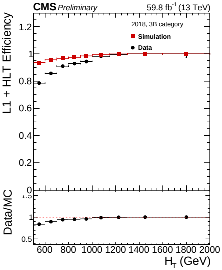
TRIGGER

Signal events are collected by the **OR** of:

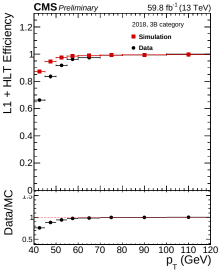
```
HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2
HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2
HLT_PFHT400_SixPFJet32_DoublePFBTagDeepCSV_2p94
HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5
HLT_PFHT430_SixPFJet40_PFBTagDeepCSV_1p5
HLT_PFHT450_SixPFJet36_PFBTagDeepCSV_1p59
HLT_PFHT1050
HLT_AK8PFJet500
HLT_AK8PFJet400_TrimMass30
HLT_AK8PFHT800_TrimMass50
```

- Trigger efficiency is measured in events with 1μ, $\geq$7 jets, $\geq$ 3 bjets
  - Reference trigger is `HLT_IsoMu24`
- 2D Scale factors are calculated to correct simulation (H_T vs $p_{\mathrm{T}\ 6\ jet}$)

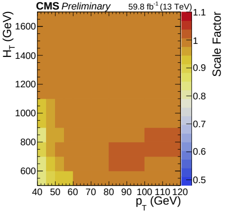
Eff vs H_T



Eff vs $p_{T,6}$



2D SF



- Syst unc applied in SF measurement

Resolved analysis

Signal region (SR):

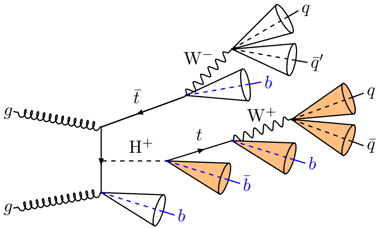
Trigger	
$\ell(\tau_h)$ veto	$p_{\mathrm{T}} > 10(20)$ GeV, $\lvert\eta\rvert < 2.4(2.3)$
≥ 7 jets	$p_T^{6th} > 40$ GeV, $p_T^{7th} > 30$ GeV, $\lvert\eta\rvert < 2.4$, Tight ID
$H_T > 500$ GeV	
≥ 3 b jets	$p_{\mathrm{T}} > 40$ GeV, DeepJet Medium WP
≥ 2 resolved top (t^{res})	$130 < m_{\mathrm{t}^{\mathrm{res}}} < 210$ GeV medium (loose) WP: 5(10)% misID rate

SR categorization based on t^{res}

- $1M1L_{t^{res}}$: medium $t^{res}_{p_{T,1}}$
  loose-not-medium $t^{res}_{p_{T,2}}$
- $2M_{t^{res}}$: both t^{res} medium tagged

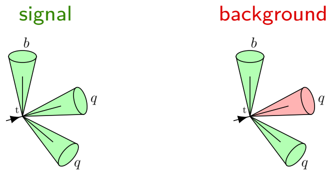
Invariant H$^{\pm}$ mass reconstruction:

$$m_{\mathrm{tb}} = t^{res}_{p_{T,1}} + \mathrm{b}_{p_{T,1}}$$



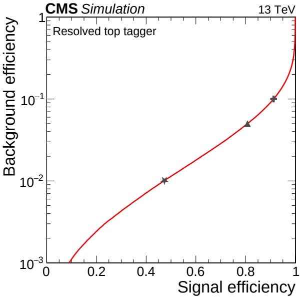
A fully connected NN is developed to reconstruct resolved top-quarks

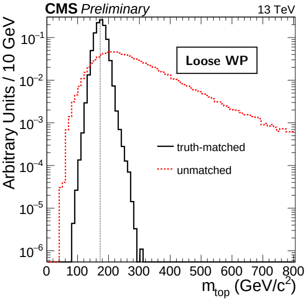
- Distinguishes trijets from top-quark decays and trijets from combinatorial background.
- Training on simulated $t\bar{t}$ events
  - Signal: truth-matched trijets
  - Background: non-matched trijets ($\geq$ 1 non-matched jet)



Mass decorrelation using sample reweighting:

- Background is reweighted such that m_{top} matches the signal.





Calibration performed

HIG-21-010 Submitted to JHEP

Documentation: AN 2021/019

Approved by JMAR group

Main background for the $H^{\pm} \to tb$ fully hadronic final state:

- QCD multijet **DATA DRIVEN**
- EWK processes (mainly $t\bar{t}$) **SIMULATION**

QCD background measurement

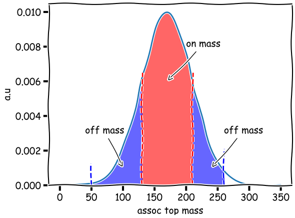
Defining 3 orthogonal control regions (CR) for each SR

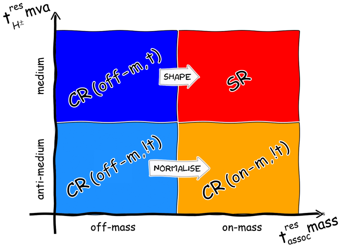
- t_{assoc}^{res} mass: On-mass $\to$ Off-mass "sidebands"
- $t_{H^{\pm}}^{res}$ mva: t-tagged (t) $\to$ non t-tagged (!t)

"ABCD" method

$$N_{QCD}^{SR} = \sum_{i}^{\text{bins}} N_{QCD,i}^{CR(off-m,t)} \cdot \left(\frac{N_{QCD,i}^{CR(on-m,!t)}}{N_{QCD,i}^{CR(off-m,!t)}} \right)$$

- $N_{\mathrm{QCD}} = N_{\mathrm{Data}} - N_{t\bar{t}} - N_{t,tt+X,EW}$
- Performed in bins of the t_{assoc}^{res} p_{T}



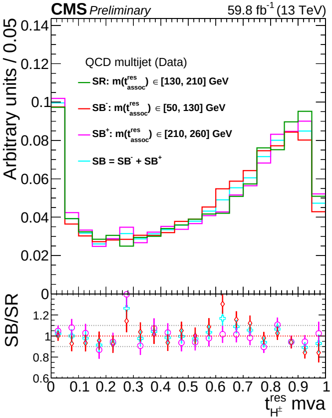


Correlation of the variables that define the ABCD method

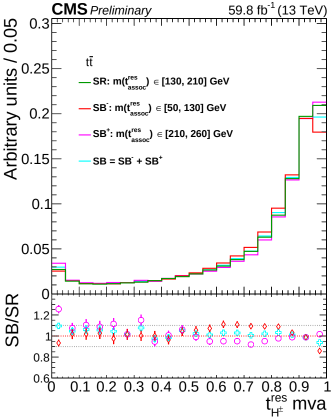
- $t^{res}_{H^{\pm}}$ mva in m(t^{res}_{assoc}) regions

$2M_{t^{res}}$

QCD (Data)



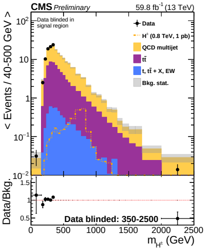
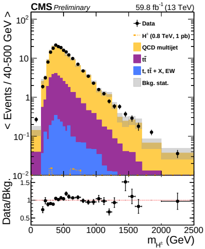
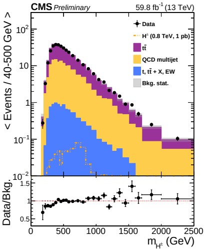
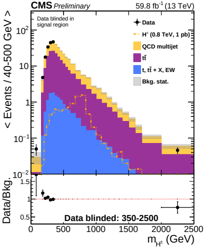
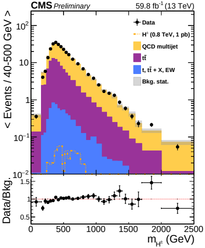
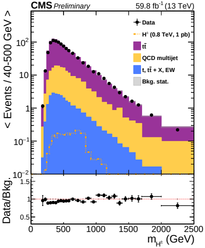
$t\bar{t}$



- No correlation between **SR** and **SB**
- Syst unc account for differences related to the SB selection

Two validation regions (VRs) for each SR

- $t\bar{t}$ enriched: $== 2$ b jets, $m_{t^{res}_{H^\pm}} \in [155, 195]$ GeV, $\Delta R_{min}(bb) > 1.2$
- QCD enriched: $== 2$ b jets, $m_{t^{res}_{H^\pm}} \notin [155, 195]$ GeV, $\Delta R_{min}(bb) < 1.0$

	SR	VR_{QCD}	$VR_{t\bar{t}}$
$1M1L_{t^{res}}$			
$2M_{t^{res}}$			

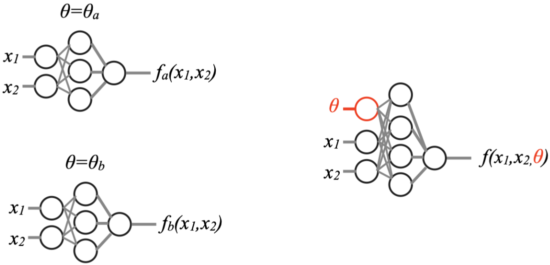
A parameterized DNN is developed to extract signal from SM background

- Signal: $H^{\pm} \rightarrow$ tb for different mass hypotheses
- Background: $t\bar{t} \rightarrow$ SR, Combinatorial $\rightarrow$ CR$^{(off-m,t)}$ $t\bar{t}$ MC

Input variables

1. $\Delta\theta(t_{H+}, b_{H+})$ in $H^{\pm}$ CM
2. $H_{T,3b}$
3. $p_T(\mathrm{bb}_{dRmin})$
4. $\mathrm{m}(\mathrm{bb}_{maxPt})$
5. $\mathrm{y23} = p^2_{T,j3}/(p_{T,j1} + p_{T,j2})^2$
6. $p_{T,b(\mathrm{H}^{\pm})}/H_{T,3b}$
7. $m_{H^{\pm}}$
8. $p_T^{Asym}(\mathrm{H}^{\pm}, \mathrm{b}_{\mathrm{H}^{\pm}})$
9. Circularity
10. Sphericity
11. Aplanarity
12. Number of medium tops
13. True mass

Parameterized DNN

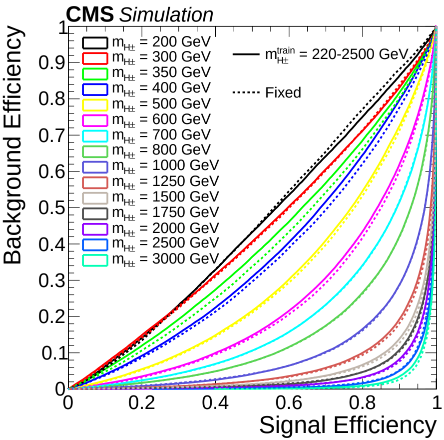


- True mass is the θ parameter
- In background events, the true mass is randomly assigned to the same values used for signal
- Training (test) is done using 2017 (2018) data

Parameterized DNN is trained using 6 different mass hypotheses.

- Training masses = [220, 350, 600, 1000, 1500, 2500] GeV **solid line**
- Performance compared to DNNs with fixed $m_{H^\pm}$ **dashed line**

ROC



- Each curve is evaluated at the true mass DNN($\mathbf{x}$,$m_{H^\pm}$)
- Comparable results!
- Good prediction even for masses not given in the training

ADD HERE THE OVERTRAINING TEST FOR 2 MASSES

Sources of systematic uncertainties

- Shape
  - Trigger efficiency
  - b (mis)tagging efficiency
  - top (mis)tagging efficiency
  - jet energy scale and resolution
  - pile-up
  - stat. unc. on the QCD transfer factors
- Constant
  - luminosity
  - top-quark mass
  - RF scale acceptance
  - cross section (scale & pdf)
  - lepton efficiency
  - syst. unc. on the QCD measurement

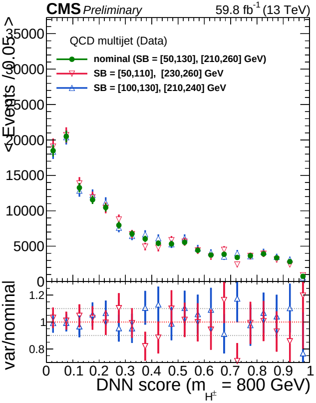
NEXT SLIDE

Estimated QCD background affected by:

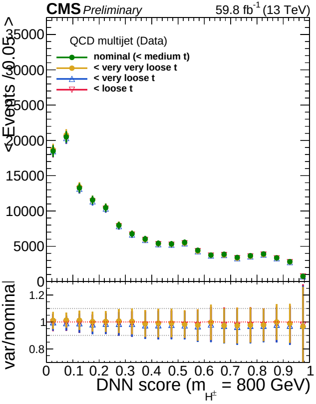
1. Sideband definition: t^{res}_{assoc} mass
2. Sideband definition: $t^{res}_{H^\pm}$ MVA
3. Subtracted background
4. Binning of the t^{res}_{assoc} p_T

$\boxed{2M_{t^{res}}}$

t^{res}_{assoc} mass



$t^{res}_{H^\pm}$ MVA

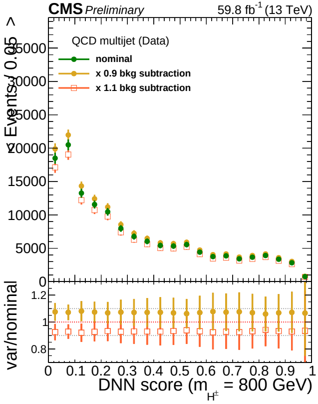


Estimated QCD background affected by:

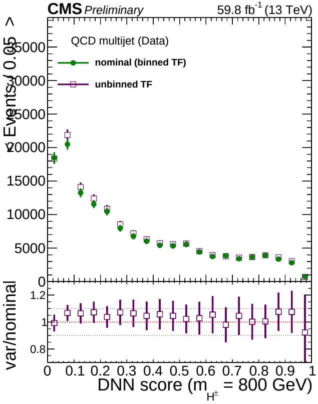
1. Sideband definition: t^{res}_{assoc} mass
2. Sideband definition: $t^{res}_{H^\pm}$ MVA
3. Subtracted background
4. Binning of the t^{res}_{assoc} p_T

$\boxed{2M_{t^{res}}}$

Subtracted bkg



t^{res}_{assoc} p_T binning

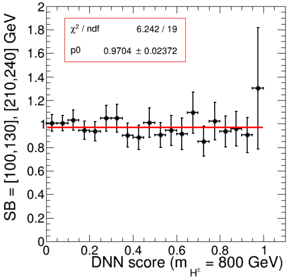


Fitted ratio to quantify the variation

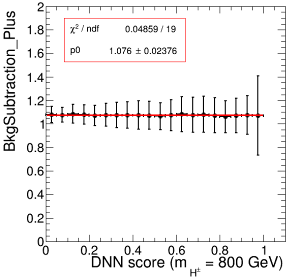
- Flat variation
- For each source:
  - Fitting the DNN score for each mass hypothesis (17 distributions)
  - Final value: distribution that gives the minimum χ^2
  - The maximum variation is applied
- All sources measured independently

$2M_{t^{res}}$

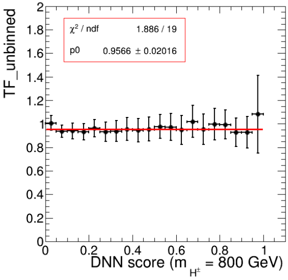
t^{res}_{assoc} mass



Subtracted bkg



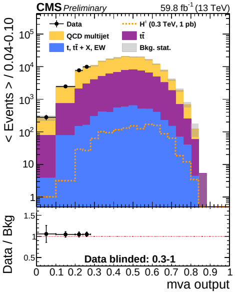
t^{res}_{assoc} p_{T} binning

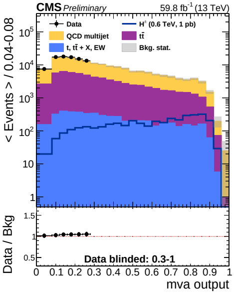


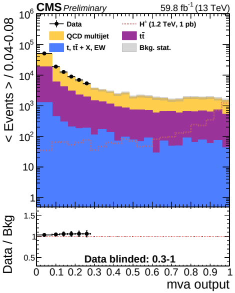
SIGNAL REGION

DNN($m_{H^\pm}$ = 300 GeV) | DNN($m_{H^\pm}$ = 600 GeV) | DNN($m_{H^\pm}$ = 1250 GeV)

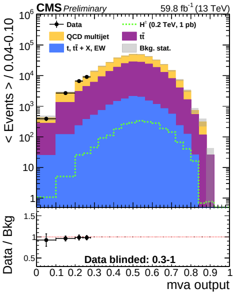
$1M1L_t^{res}$

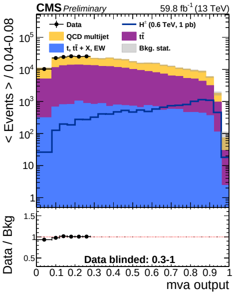


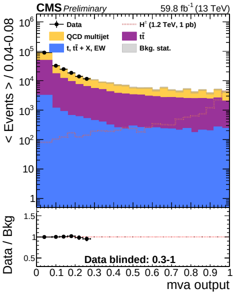




$2M_t^{res}$



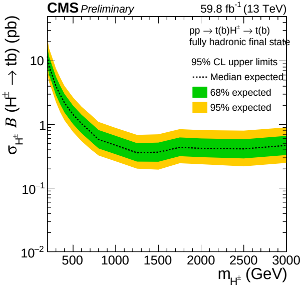




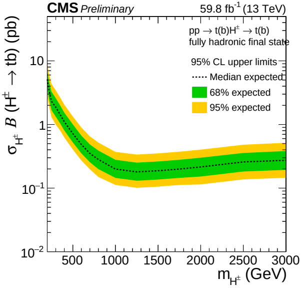
Expected limits on $\sigma_{H^{\pm}t(b)} \times \mathcal{B}(H^{\pm} \to tb)$

Preliminary

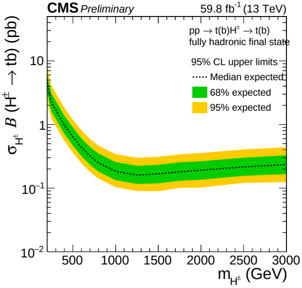
$1M1L_t$



$2M_t$



Combined



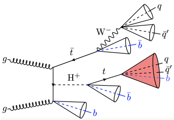
- Statistical uncertainties only
- Sensitivity comes to a plateau for $m_{H^{\pm}} > 1250$ GeV
- Results improved by a factor of 2 wrt 2016 data analysis

Boosted analysis

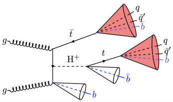
Two SRs based on the number of boosted tops (t^{bst}):

SR1 : $N_{t^{bst}} == 1$
SR2 : $N_{t^{bst}} == 2$

SR1



SR2

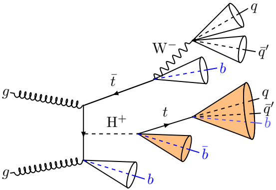


Preliminary

SR1	SR2	
Trigger	Trigger	
ℓ veto	ℓ veto	same as resolved
$= 1\ t^{bst}$	$= 2\ t^{bst}$	$p_{\mathrm{T}} > 400$ GeV, $\lvert\eta\rvert < 2.4$, `PNet_TvsQCD` Medium WP (5% misID rate)
≥ 4 jets	≥ 2 jets	$p_{\mathrm{T}} > 40$ GeV, $\lvert\eta\rvert < 2.4$, tight ID, $H_T > 500$ GeV
≥ 2 b jets	≥ 1 b jets	DeepJet Medium WP
$\leq 2\ t^{res}$		custom DNN loose, $130 < m_{\mathrm{t}^{\mathrm{res}}} < 210$ GeV
$\Delta R(t^{bst}, b^{ldg}) > 1.2$		
$\max(m_{bb}) > 200$ GeV		

Invariant H$^{\pm}$ mass reconstruction:

$$m_{\mathrm{tb}} = t^{bst}_{p_{T,1}} + \mathrm{b}_{p_{T,1}}$$

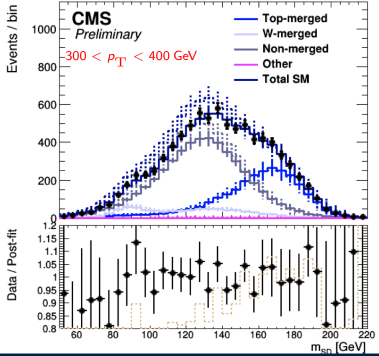


Boosted top/W jets identified with the `ParticleNet` discriminators

- Calibration performed in semileptonic $t\bar{t}$ events (tag & probe)
- 3 jet types: top-matched, W-matched, non-matched
- 2D (m_{SD}^{jet} , jet p_{T}) templates derived for each:
  - jet type
  - WP of the tagger
  - Pass/Fail the selected WP

Efficiency in Data: Simultaneous fit of all jet types, for both pass/fail events

Pass, TvsQCD, 2018, misID=1%



- $SF_i = \frac{\epsilon_{Data,i}}{\epsilon_{MC,i}}$, i = top, W, non-matched
- SF_i free parameters, $\epsilon_{MC,i}$ contant
- Number of pass/fail events from each jet-type category in data determined by the SF_i ($\epsilon_{Data,i} = \frac{P_i}{P_i+F_i}$)
- Scale factors expressed vs jet p_{T}
- Presentation in JME

JET TAGGING

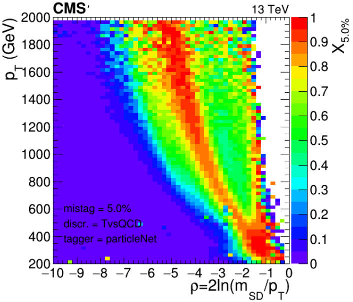
Boosted top jets t^{bst} identification with `ParticleNet_TvsQCD`

Designed decorrelated tagger (DDT)

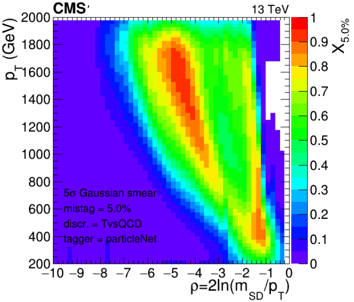
- A 3D map of the tagger's score for a fixed mID rate vs p_{T} and $\rho = \ln(m_{SD}^2/p_T^2)$

Preliminary

before smearing



5σ Gaussian smear



- Calculated with simulation QCD multijet events
- For each (p_{T}, ρ) bin: estimate the WP that corresponds to 5% mID rate: $X(5\%)$
- Transformed score: $X(DDT) = X_{raw}$ - $X(5\%)$ **p_t, ρ dependent**
- Selection requirement $X(DDT) > 0$

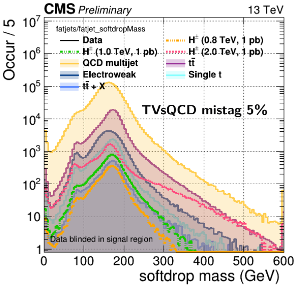
Main background:

- $t\bar{t}$ (merged-t, merged-W, non-merged)
- QCD multijet
- other (minor)

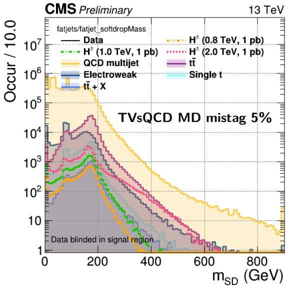
- 2D (m^t_{SD}, m_{tb}) templates derived from MC simulation
- Signal extraction: Signal, background simultaneous fit of (m^t_{SD}, m_{tb})

Preliminary

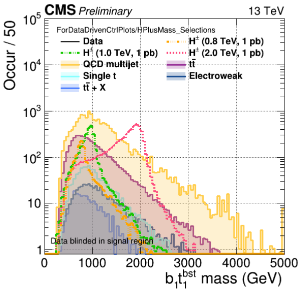
m^t_{SD}



m^t_{SD}



m_{tb}



Summary

SUMMARY

Search for $\mathrm{H}^{\pm} \rightarrow$ tb in fully hadronic final state presented with 2018 UL Data

New with respect to the previous results:

- Resolved Analysis:
  - Top tagging: custom mass-decorrelated DNN (almost published!)
  - Event categorization based on the number of medium tagged t^{res}
  - Very good data-driven QCD background prediction
  - Mass parameterized DNN score used as a signal discriminant
  - Preliminary expected limits using 2018 data with statistical uncertainties only
- Boosted Analysis:
  - boosted top indentification with `ParticleNet` (mva-based)
  - New category with 1 boosted and 1 resolved top
  - Designed decorrelated top tagger to eliminate mass sculpting effects

FUTURE WORK

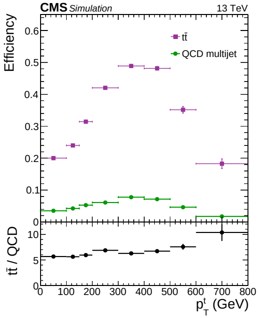
- Resolved Analysis:
  - Incorporate the systematic uncertainties IN PROGRESS
  - Final touches on the parameterized DNN IN PROGRESS
- Boosted Analysis:
  - `ParticleNet` W/t re-calibration (L.Paizanos) IN PROGRESS
  - Study the merged-W category IN PROGRESS
  - Categorization based on the top tagging rate and N_{bjets}^{extra} IN PROGRESS
  - Extract QCD and $\mathrm{t\bar{t}}$ templates for the fit IN PROGRESS
  - Produce first limits with simultaneous 2D-fit in $(m_{\mathrm{SD}}^{j}, m_{\mathrm{tb}})$ plane
  - Address systematic uncertainties
- Finalize and release documentation
- Complete the analysis with entire Run II (target Moriond23)

BACKUP

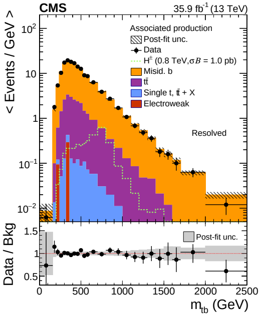
Resolved CADI: HIG-18-015

- Resolved t (t^{res}) identification: custom top tagger (BDT)
- Selected events contain ≥ 7 jets, ≥ 3 b-tagged, 2 t^{res}
- $H^{\pm}$mass reconstruction (m_{bt}): leading p_{T} t^{res} + leading p_{T} b jet
- Main background:
  - Misid. B: From data using CRs (ABCD method)
  - Genuine B: from simulation
- m_{bt} is used to extract the signal in the presence of the SM background.

t^{res} efficiency



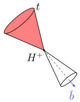
m_{bt} Post-fit



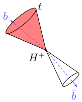
Boosted CADI: HIG-18-015

- Events are split in four main categories

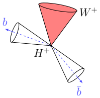
t0b



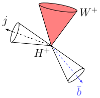
t1b



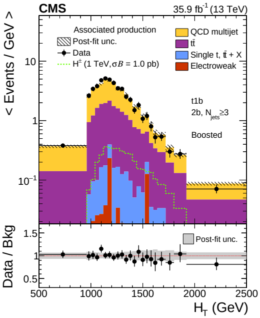
Wbb



Wbj



- Boosted t/W identification:
  - Based on m_{SD}, τ_N, N_b subjets
- Further categorization according to:
  - $N_b \in [=1, =2, \geq 3]$
  - $N_j^{extra} \in [<3, \geq 3]$
  - $m_{\text{tb}} \in$ [below, in, above] of FWHM of signal



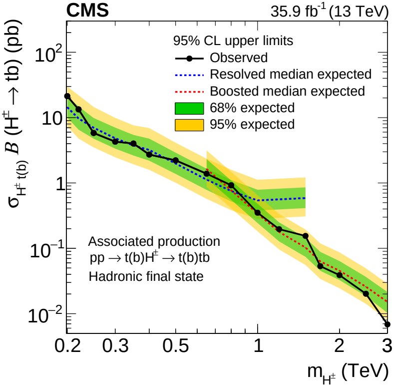
- Main background

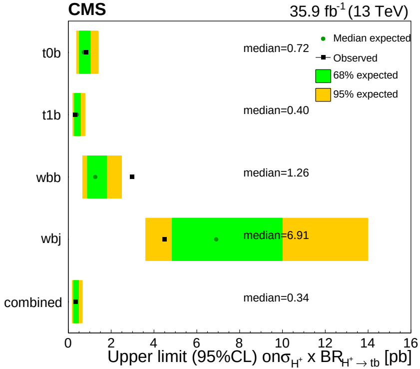
QCD : from data using CRs (inverted τ_N), sidebands with $m_{\text{tb}} \in$ [below, above])

$t\bar{t}$: from sim., normalized in CR with 1 ℓ

- H_T is used to extract the signal from SM background inside the m_{bt} window.

Upper limits on $\sigma_{\mathrm{H}^{\pm}t(b)} \times \mathcal{B}(\mathrm{H}^{\pm} \rightarrow \mathrm{tb})$





$m_{H^{\pm}}$ = 1000 [GeV]

- Resolved and Boosted overlayed limits
- No excess above the estimated background
- Interpretation in hMSSM: max. $\tan\beta = 0.88$ excluded for $m_{H^{\pm}} = 0.20$-0.55 TeV

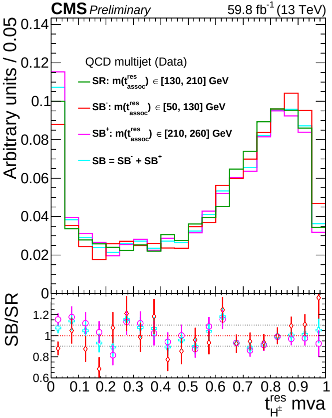
- Boosted analysis categories
- Most sensitive category is *t1b*
- Least sensitive category is *Wbj*

Correlation of the variables that define the ABCD method

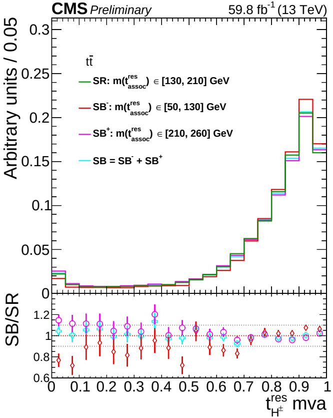
- $t^{res}_{H^\pm}$ mva in m(t^{res}_{assoc}) regions

$1M1L_{t^{res}}$

QCD (Data)



$t\bar{t}$



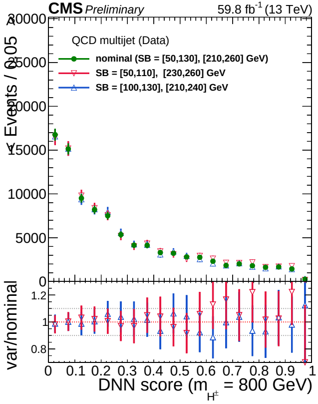
- No correlation between **SR** and **SB**
- Syst unc account for differences related to the SB selection

Estimated QCD background affected by:

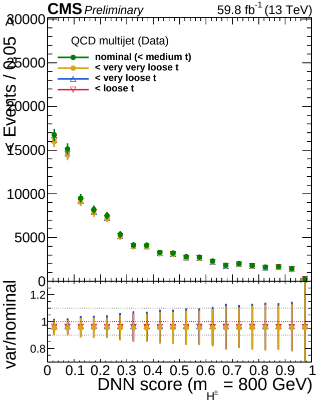
1. Sideband definition: t^{res}_{assoc} mass
2. Sideband definition: $t^{res}_{H^\pm}$ MVA
3. Subtracted background
4. Binning of the t^{res}_{assoc} p_{T}

$\boxed{1M1L_{t^{res}}}$

t^{res}_{assoc} mass



$t^{res}_{H^\pm}$ MVA

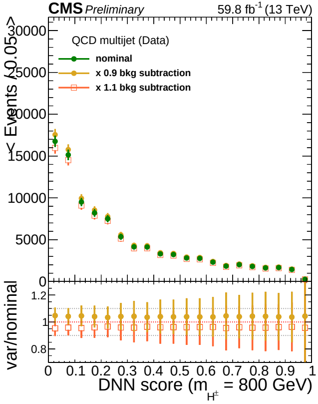


Estimated QCD background affected by:

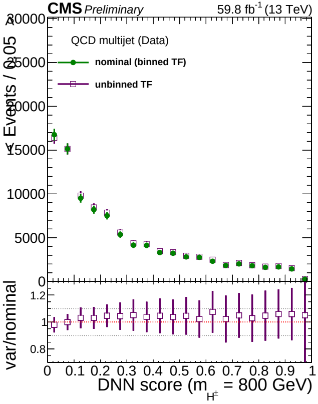
1. Sideband definition: t^{res}_{assoc} mass
2. Sideband definition: $t^{res}_{H^{\pm}}$ MVA
3. Subtracted background
4. Binning of the t^{res}_{assoc} p_{T}

$1M1L_{t^{res}}$

Subtracted bkg



t^{res}_{assoc} p_{T} binning



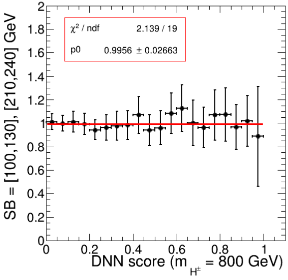
SYSTEMATIC UNCERTAINTIES QCD measurement

Fitted ratio to quantify the variation

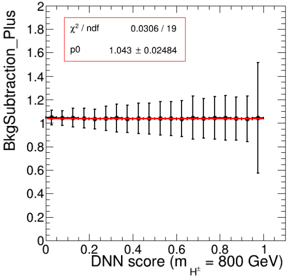
- Flat variation
- For each source:
  - Fitting the DNN score for each mass hypothesis (17 distributions)
  - Final value: distribution that gives the minimum χ^2
  - The maximum variation is applied
- All sources measured independently

$1M1L_t^{res}$

t_{assoc}^{res} mass



Subtracted bkg



t_{assoc}^{res} p_{T} binning