Search for charged Higgs boson with the $H^{\pm} \rightarrow tb$ 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]

[1]UCY [2]HIP

Friday 25th November, 2022

Many BSM theories need to enlarge their Higgs sector to two Higgs doublets

- The minimal two-Higgs-doublet models (2HDMs) predict 5 physical states:
  - two neutral, $\mathcal{CP}$-even particles h and H ($m_{\mathrm{h}} \leq m_{\mathrm{H}}$)
  - one neutral, $\mathcal{CP}$-odd particle A^0
  - two charged Higgs bosons $\mathrm{H}^{\pm}$

SM fermion coupling to 2HDs (no FCNCs):

I All quarks & leptons couple to Φ_2

II All u-type to Φ_2 and all d-type & ℓ to Φ_1

X Both u & d types couple to Φ_2, all ℓ to Φ_1

Y Roles of two doublets reversed wrt type II

Type	u	d	ℓ
I	Φ_2	Φ_2	Φ_2
II	Φ_2	Φ_1	Φ_1
III (X)	Φ_2	Φ_2	Φ_1
IV (Y)	Φ_2	Φ_1	Φ_2

For each 2HDMs type there are 7 free parameters (incl. m_{h}, m_{H}, m_{A}, $m_{\mathrm{H}^{\pm}}$):

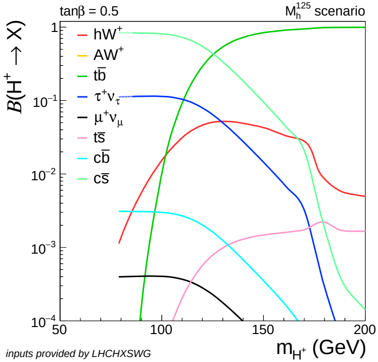
5. $\tan\beta \equiv \frac{v_2}{v_1}$, the ratio of the Higgs doublet VEVs
6. $\sin(\beta - \alpha)$, α: the mixing angle of the $\mathcal{CP}$-even states
7. m_{12}, diagonal term of the mass matrix of the Higgs doublets

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

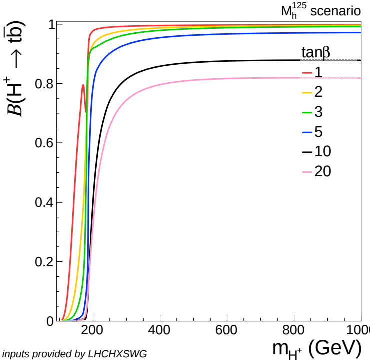
- Light $m_{H^{\pm}} < m_t - m_b$, intermediate $m_{H^{\pm}} \sim m_t$, heavy $m_{H^{\pm}} > m_t + m_b$

Decay BRs model-dependent $\Rightarrow$ different searches constrain different scenarios.

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



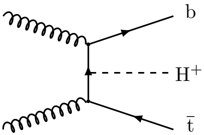
M_h^{125}, $H^{\pm} \to tb$



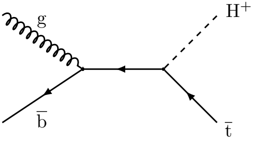
BRs of $H^{\pm} \to tb$ dominanates at high $m_{H^{\pm}}$, for wide range of $\tan\beta$

This analysis searches for a heavy $H^{\pm}$

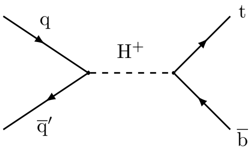
4FS



5FS

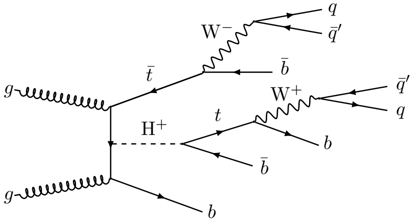


s-channel



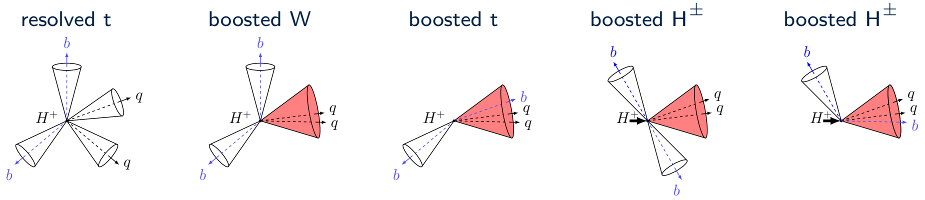
Fully-hadronic final state of associated production characterised by:

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



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

- **Resolved t**: At moderate $m_{\mathrm{H}^\pm}$ & $p_{\mathrm{T,H}^\pm}$ 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
- **Boosted $\mathrm{H}^\pm$**: As $p_{\mathrm{T,H}^\pm}$ increases its decay products become collinear



Previous results

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

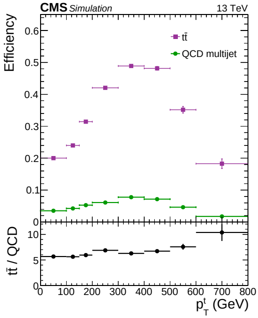
This work

- **Resolved t**, **Boosted t**
- Full Run II data
- This talk: status of 2018 data
- Last report (HExtended): 25 Oct 2021

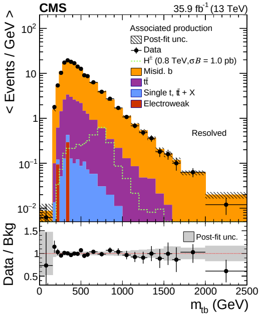
Resolved CADI: HIG-18-015

- 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



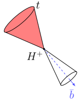
m_{bt} Post-fit



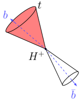
Boosted CADI: HIG-18-015

- Events are split in four main categories

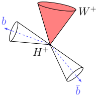
t0b



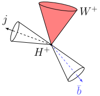
t1b



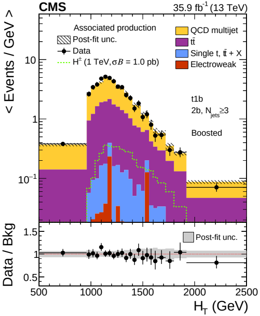
Wbb



Wbj



- Boosted t/W identification: Based on m_{SD}, τ_N, $N_{b\ \mathrm{subjets}}$
- H$^{\pm}$mass reconstruction (m_{bt}): t + leading p_{T} b jet
- Further categorization according to:
  - $N_b \in [=1, =2, \geq 3]$
  - $N_j^{extra} \in [< 3, \geq 3]$
  - $m_{\mathrm{tb}} \in$ [below, in, above] of FWHM of signal



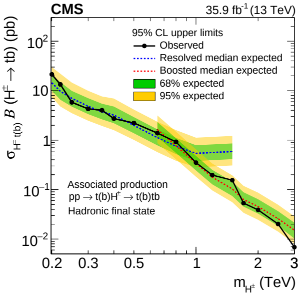
- Main background

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

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

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

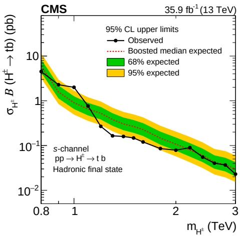
Associated production



Resolved and Boosted analysis overlayed limits

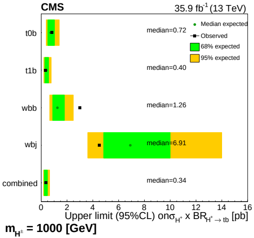
- Resolved: most stringent limits at $m_{H^{\pm}} \leq 0.8$ TeV
- Reported limit at each $m_{H^{\pm}}$: analysis with best expected sensitivity.
- No excess above the estimated background
- Interpretation in hMSSM: maximum $\tan\beta = 0.88$ is excluded for $m_{H^{\pm}} = 0.20$-0.55 TeV

s-channel



Boosted categories

Associated production



Boosted analysis:

- Upper limits in the s-channel production for $m_{H^{\pm}}$ 0.8 to 3 TeV
- Most sensitive main category is $t1b$
- Least sensitive category is Wbj

Three main categories with different topology

1. Both top candidates are resolved RESOLVED
2. Associated top is resolved, top from $H^{\pm}$ is boosted BOOSTED
3. Both top candidates are boosted BOOSTED

- Further categorization based on the top tagging rate (high/low purity)

- This analysis targets full Run II data
- This talk presents the status of 2018 data (RunIISummer20UL18)

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

Signal region (SR):

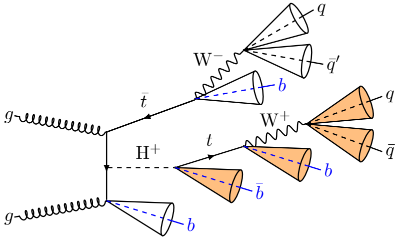
Trigger	H_T + multijet + 1 or 2 b jets
e veto	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.4$, Loose miniIso, cutBasedElectronID (veto)
μ veto	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.4$, Loose miniIso isCutBasedIDLoose
τ veto	$p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.3$, DeepTau $D_{\mathrm{e}}^{\mathrm{vloose}}$, $D_{\mu}^{\mathrm{medium}}$, D_{j}^{loose}
≥ 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
≥ 1 resolved top (t^{res})	custom DNN medium, $130 < m_{\mathrm{t}^{\mathrm{res}}} < 210$ GeV

SR categorization based on t^{res}

- $1M1L_{t^{res}}$: medium $t^{res}_{\mathrm{H}^{\pm}}$
  loose-not-medium t^{res}_{assoc}
- $2M_{t^{res}}$: both t^{res} medium tagged

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

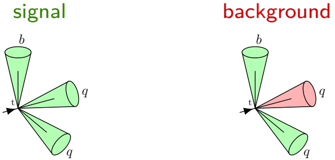
$$m_{\mathrm{tb}} = t^{res}_{ldg\ p_T} + \mathrm{bjet}^{free}_{ldg\ p_T}$$



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

- Distinguishes trijets from top-quark decays and trijets from combinatorial background.

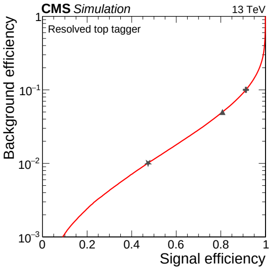
- Training on simulated $t\bar{t}$ events
  - Signal: truth-matched trijets
  - Background: non-matched trijets

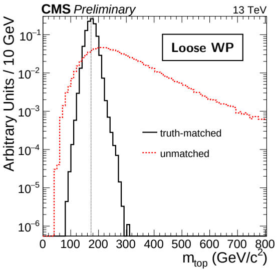


Mass decorrelation using sample reweighting:

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

SF vs t^{res} p_{T} measured in a region with 1 isolated ℓ





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

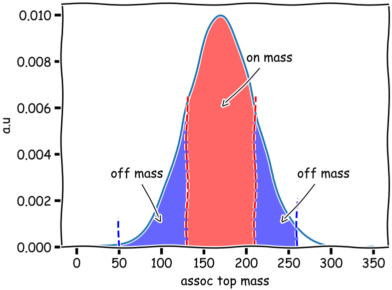
- t^{res}_{assoc} mass: On-mass → Off-mass "sidebands"
- $t^{res}_{H^{\pm}}$ mva: t-tagged (t) → non t-tagged (!t)

"ABCD" method

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

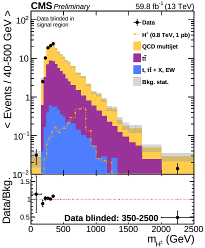
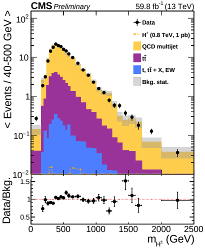
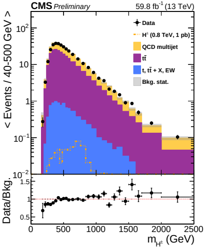
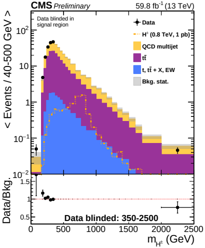
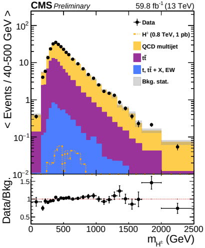
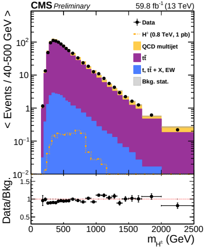
- Performed in bins of the t^{res}_{assoc} p_{T}:
  - $2M_{t^{res}}$: $p_{\mathrm{T}} \in [0, 100, 300, \infty]$ GeV
  - $1M1L_{t^{res}}$: $p_{\mathrm{T}} \in [0, 175, \infty]$ GeV

Sidebands



Two validation regions (VRs) for each SR

- $t\bar{t}$ enriched: $== 2$ b jets, $m_{t^{res}_{H^{\pm}}} \in [145, 195]$ GeV, $\Delta R_{min}(bb) > 1.2$
- QCD enriched: $== 2$ b jets, $m_{t^{res}_{H^{\pm}}} \notin [145, 195]$ GeV, $\Delta R_{min}(bb) < 0.9$

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

RESOLVED signal extraction

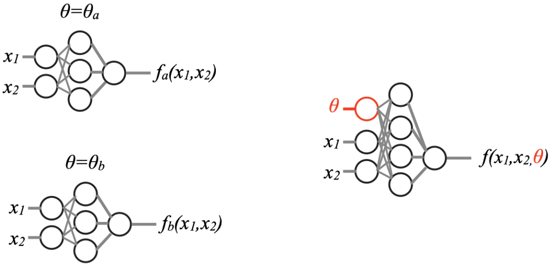
A parameterized DNN is developed to extract signal from SM background

- 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

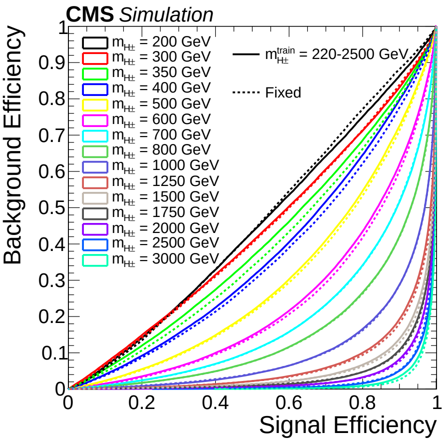


- True mass is the θ parameter
- In background events, the true mass is randomly assigned to the same values used for signal
- Training (test) is done using 2017 (2018) data

Parameterized DNN is trained using 6 different mass hypotheses.

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

ROC



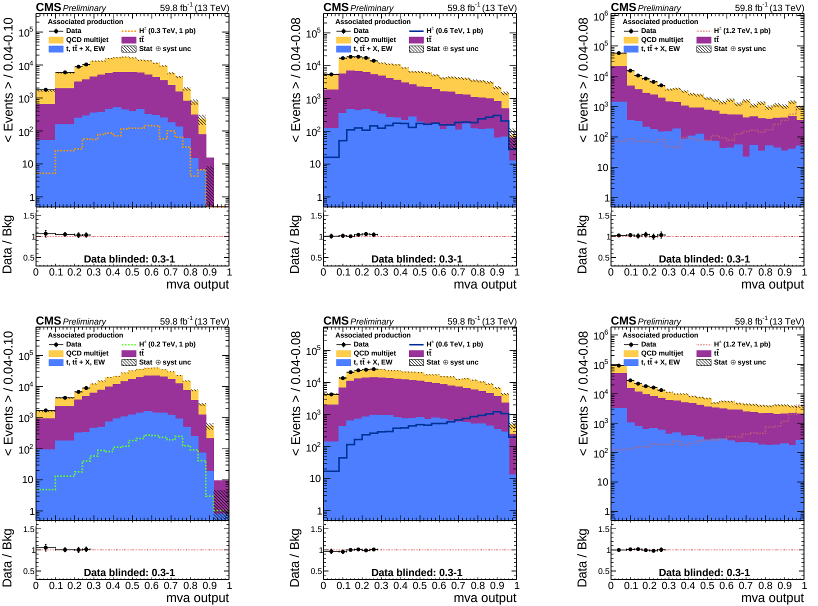
- Each curve is evaluated at the true mass DNN($\mathbf{x}, m_{H^\pm}$)
- Comparable results!
- Good prediction even for mass values that were not seen during the training

RESOLVED signal region

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

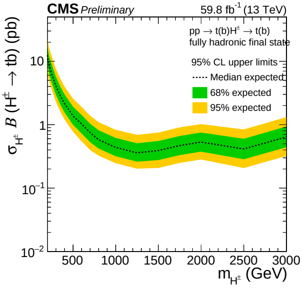
$1M1L_t^{res}$

$2M_t^{res}$

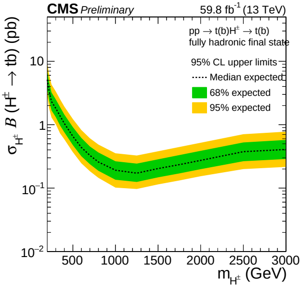


Expected limits on $\sigma_{\mathrm{H}^\pm t(b)} \times \mathcal{B}(\mathrm{H}^\pm \to \mathrm{tb})$

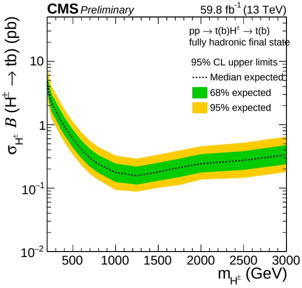
$1\mathrm{M}1\mathrm{L}_\mathrm{t}$



$2\mathrm{M}_\mathrm{t}$



Combined

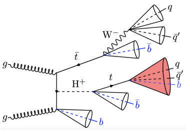


- Statistical uncertainties only
- Sensitivity comes to a plateau for $m_{H^\pm} > 1000$ GeV
- Results improved by a factor of 2 wrt 2016 data analysis **FIXME**

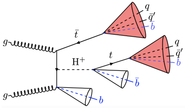
Two SRs based on the number of boosted tops (t^{bst}):

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

SR1



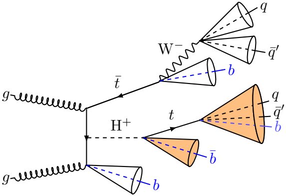
SR2



SR1	SR2	
Trigger	Trigger	H_T + multijet + 1 or 2 b jets H_T + AK8 jet + trim mass
ℓ veto	ℓ veto	same as resolved
$= 1\ t^{bst}$	$= 2\ t^{bst}$	$p_T > 400$ GeV, $\lvert\eta\rvert < 2.4$, `ParticleNet_TvsQCD` Medium WP
≥ 4 jets	≥ 2 jets	$p_T > 40$ GeV, $\lvert\eta\rvert < 2.4$, tight ID, $H_T > 500$ GeV
≥ 2 b jets	≥ 1 b jets	DeepJet Medium WP
$\leq 2\ t^{res}$	$\geq 0\ t^{res}$	custom DNN loose, $130 < m_{t^{res}} < 210$ GeV
$\Delta R(t^{bst}, b^{ldg}) > 1.2$	$\Delta R(t^{bst}, b^{ldg}) > 0.0$	
$\max(m_{bb}) > 200$ GeV	$\max(m_{bb}) > 0$ GeV	

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

$m_{tb} = t^{bst} + \text{bjet}^{ldg\ p_T}$

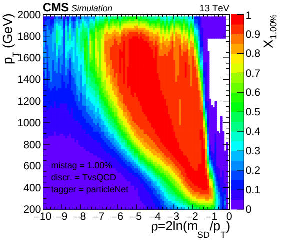


Boosted top jets t^{bst} are identified using the `ParticleNet_TvsQCD` discriminant

Designed decorrelated tagger (DDT)

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

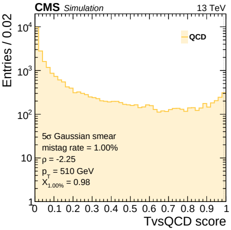
3D map

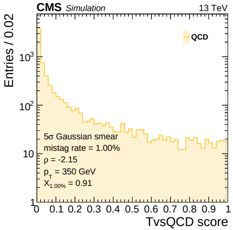


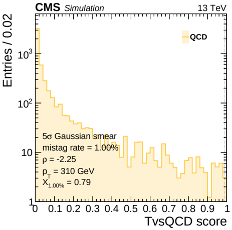
- For each (p_{T}, ρ) bin: estimate the WP that corresponds to %y mID rate: $X(y\%)$ QCD MC
- Transformed variable: p_t, ρ **dependent**

  $X(DDT) = X$ - $X(y\%)$
- Selection requirement $X(DDT) < 0$

$X(1\%)$ in (p_{T}, ρ) bins:







Main background: QCD multijet, $\mathrm{t\bar{t}}$

- 2D (m^{t}_{SD}, m_{tb}) templates derived from MC simulation
- Normalization: from sidebands on m_{tb} or CRs with inverted requirements FIXME
- ...

SUMMARY

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

New with respect to the previous results:

- Three search topologies containing resolved and/or boosted tops
- Resolved Analysis:
  - Top tagging: custom mass-decorrelated DNN (almost published!)
  - Event categorization based on the number of medium tagged t^{res}
  - QCD background measurement shows good agreement in validation region
  - Mass parameterized event-based tagger used as a final discriminant
  - First expected limits with 2018 data with statistical uncertainties only
- Boosted Analysis:
  - boosted top indentification with `ParticleNet` (mva-based)
  - Designed decorrelated t-tagger eliminates the mass sculpting effect
  - Event categorization based on the number of t^{bst}
  - First results show improved signal sensitivity and significance FIXME

FUTURE WORK

- Resolved Analysis:
  - Incorporate the systematic uncertainties IN PROGRESS
  - Final touches on the event-based tagger IN PROGRESS
- Boosted Analysis:
  - Finalize background data driven method IN PROGRESS
  - Produce first expected limits
- Finalize and release documentation
- Extend the analysis with entire Run II

Thank you!

BACKUP