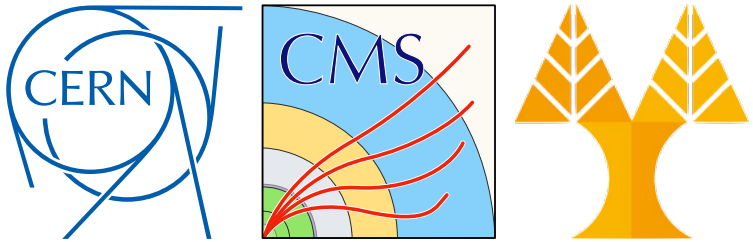
QCD background data-driven estimation

UCY-CMS Group Weekly Meeting

S. Konstantinou

University of Cyprus (UCY)

Sunday 13th 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 ($m_h \leq m_H$)
  - one neutral, $\mathcal{CP}$-odd particle A^0
  - two charged Higgs bosons $H^{\pm}$

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, $m_{H^{\pm}}$):

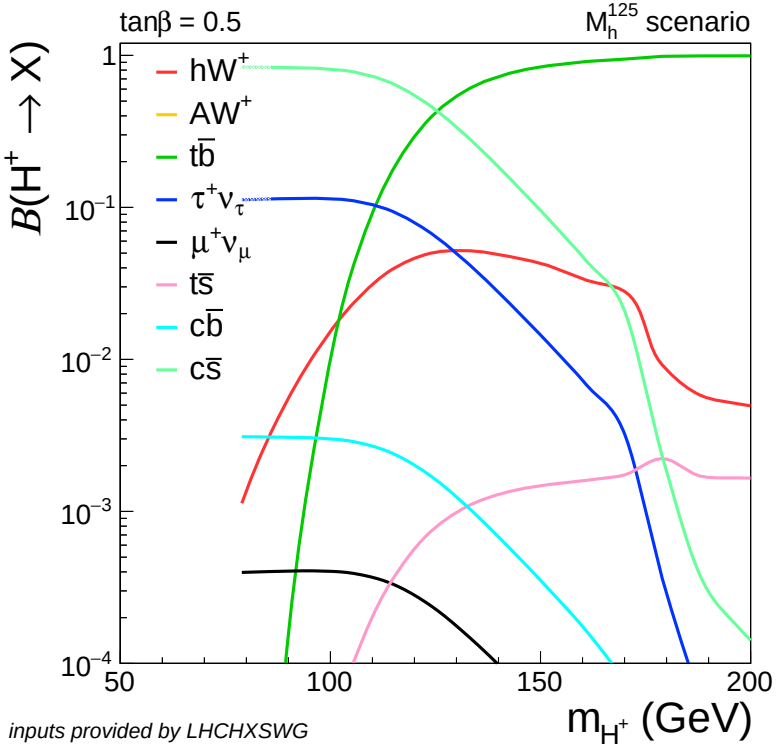
5. $\tan\beta \equiv \frac{v_2}{v_1}$, the ratio of the Higgs doublet VEVs
6. $\sin(\beta - \alpha)$, α: the mixing angle of the $\mathcal{CP}$-even states
7. m_{12}, diagonal term of the mass matrix of the Higgs doublets

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

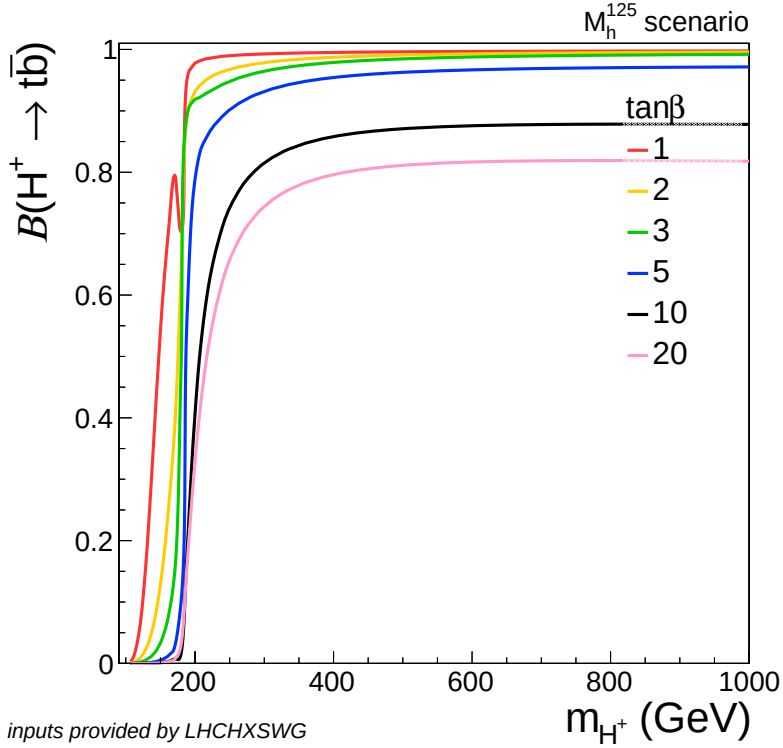
- Light $m_{H^{\pm}} < m_t - m_b$, intermediate $m_{H^{\pm}} \sim m_t$, heavy $m_{H^{\pm}} > m_t + m_b$

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

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



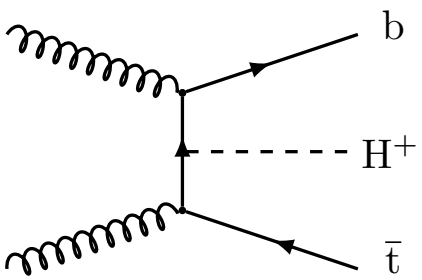
M_h^{125}, $H^{\pm} \to tb$



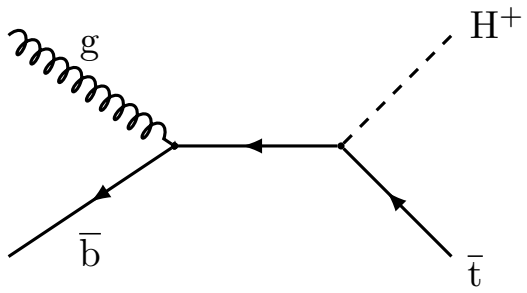
BRs of $H^{\pm} \to tb$ dominanates at high $m_{H^{\pm}}$, for wide range of tan β

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

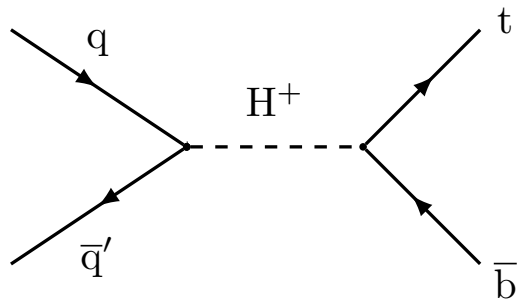
4FS



5FS

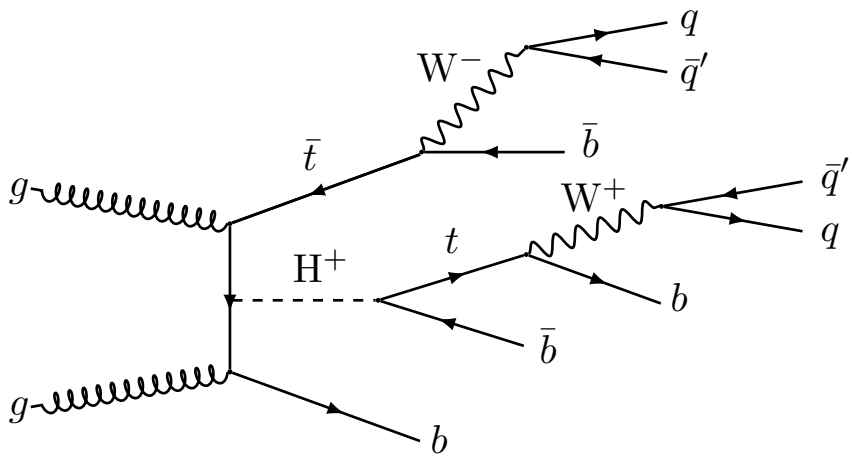


s-channel



Fully-hadronic final state of associated production characterised by:

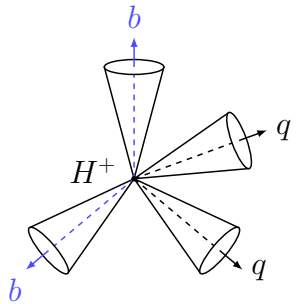
- ▶ High jet & b jet multiplicities
- ✓ Large branching ratio $\mathcal{B} \simeq 46\%$
- ✓ Invariant mass reconstruction of $H^{\pm}$
- ✗ QCD multijet & $t\bar{t}$ background
- ✗ Combinatorial (self-)background



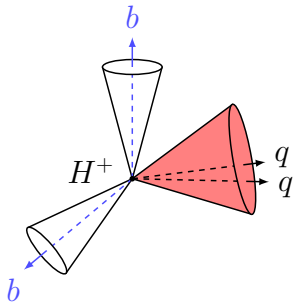
Various $\mathrm{m}_{\mathrm{H}^{\pm}}$ reconstruction techniques available due to signal process kinematics:

- **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

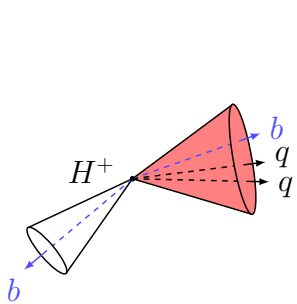
resolved t



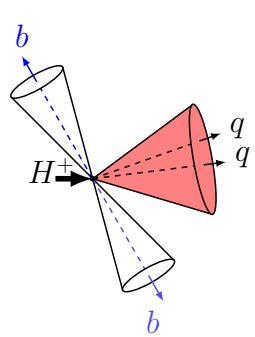
boosted W



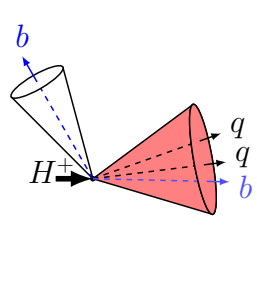
boosted t



boosted $\mathrm{H}^{\pm}$



boosted $\mathrm{H}^{\pm}$



Previous results

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

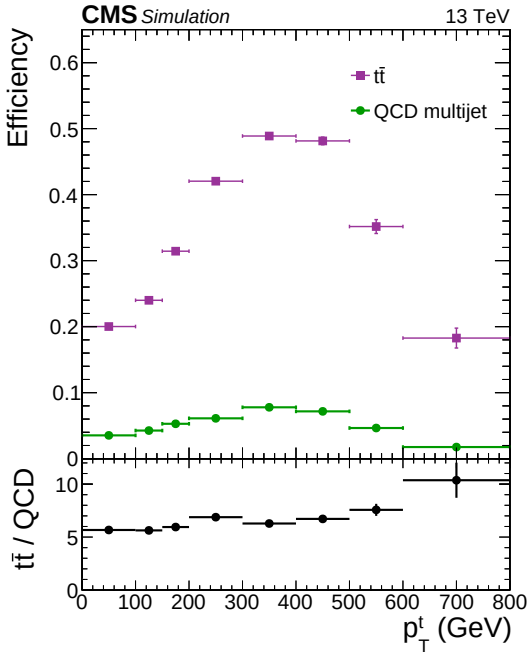
This work

- **Resolved t**, **Boosted t**
- Full Run II data
- This talk: status of 2017-2018 data
- Last report (HExtended): 25 Oct 2021

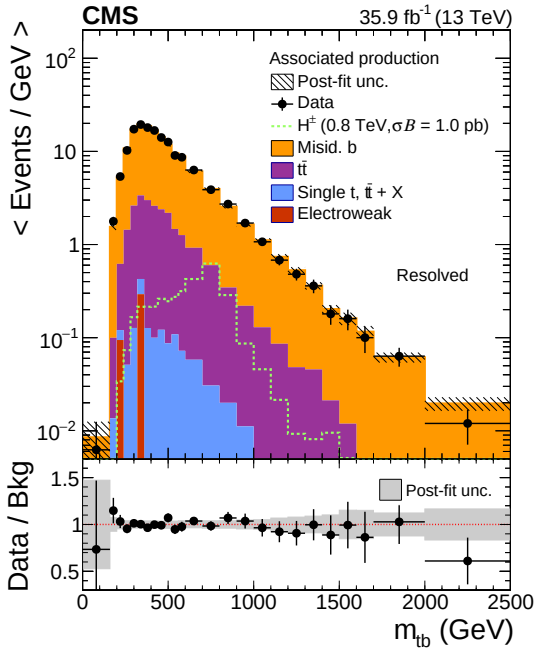
Resolved (UCY, HIP)

- Resolved t (t^{res}) identification: custom top tagger (BDT)
- Selected events contain $\geq$ 7 jets, $\geq$ 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



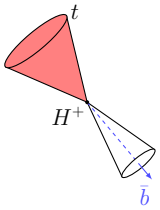
m_{bt} Post-fit



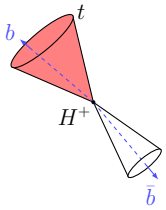
Boosted (MIT, BUAP)

- Events are split in four main categories

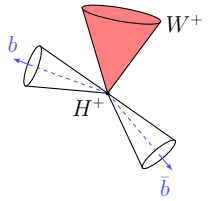
t0b



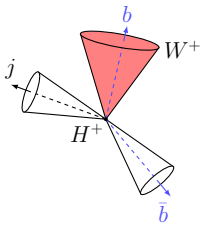
t1b



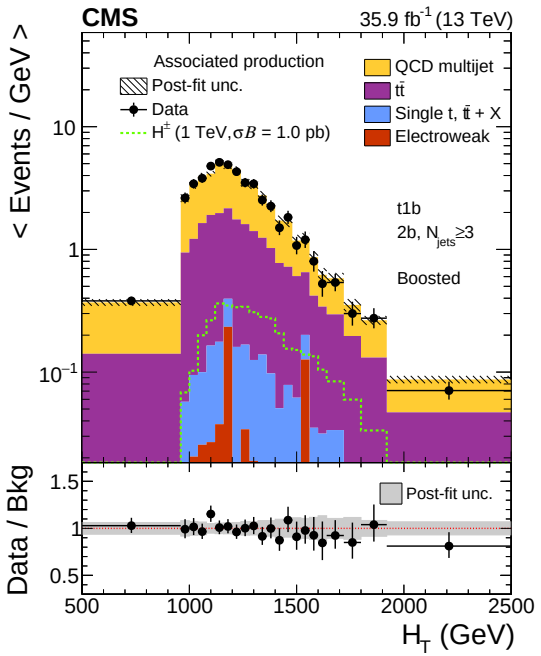
Wbb



Wbj



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



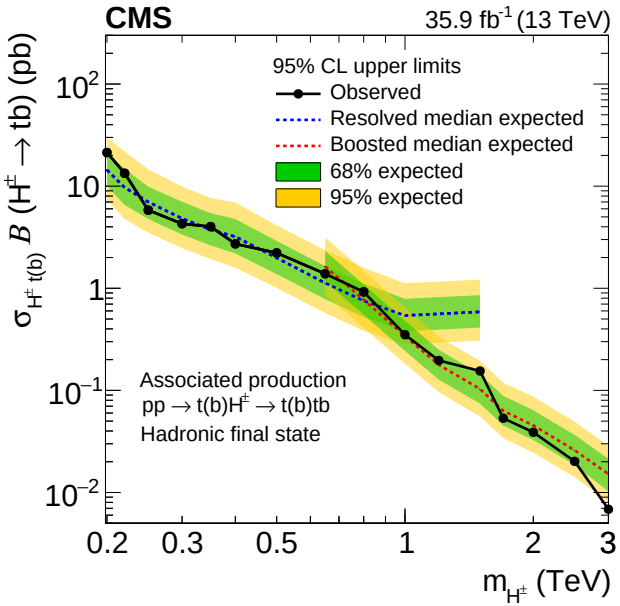
- Main background

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.

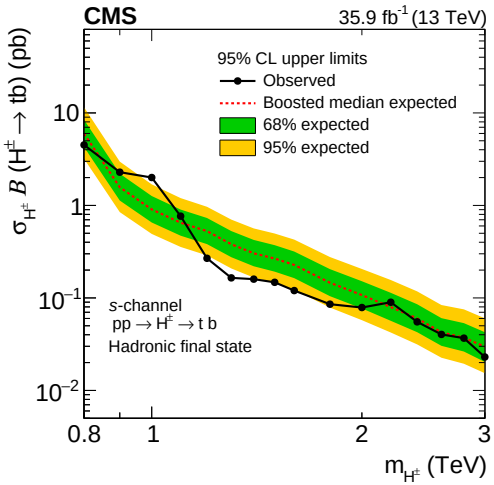
Associated production



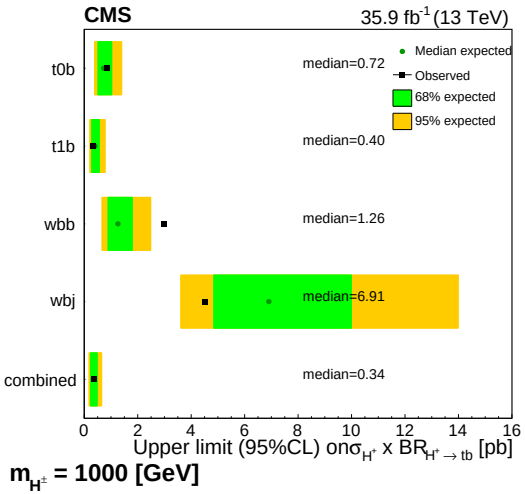
Resolved and Boosted analysis overlayed limits

- Boosted: best sensitivity for $m_{H^\pm} > 0.8$ TeV
- Reported limit at each mass value is determined by the analysis with the best expected sensitivity.
- 21.3 to 0.007 pb for masses 0.2 to 3 TeV
- 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



Assoc. boosted categories



- Boosted analysis sets upper limits in the s-channel production
  - 4.5-0.023 pb for $m_{H^\pm}$ 0.8 to 3 TeV
- Boosted categories
  - Most sensitive main category is *t1b*
  - Least sensitive category is *Wbj*

Signal region (SR):

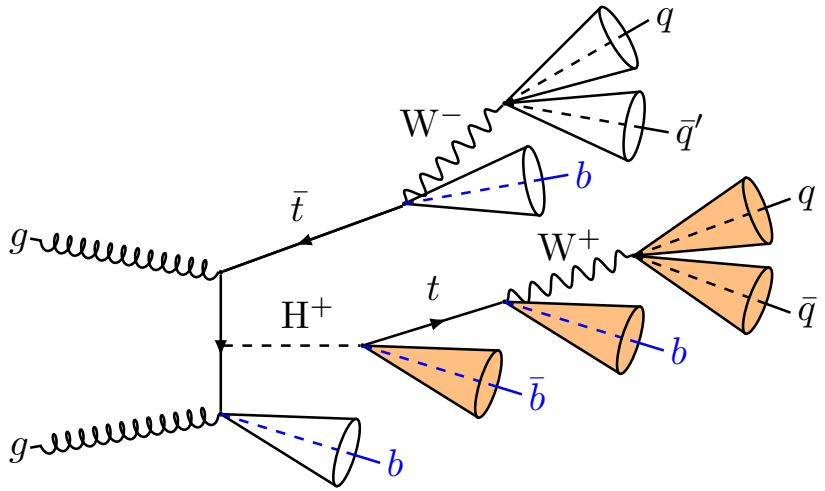
Trigger	H_T + multijet + 1 or 2 b jets
e veto	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.4$, Loose miniIso, cutBasedElectronID (veto)
μ veto	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.4$, Loose miniIso isCutBasedIDLoose
τ veto	$p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.3$, DeepTau $D_{\mathrm{e}}^{\mathrm{vloose}}$, $D_{\mu}^{\mathrm{medium}}$, 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:

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

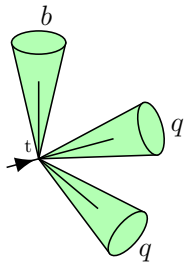


RESOLVED top tagging

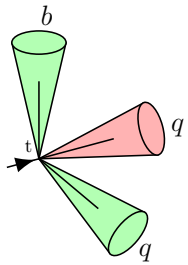
A fully connected NN is developed to reconstruct resolved top-quarks

- 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

signal



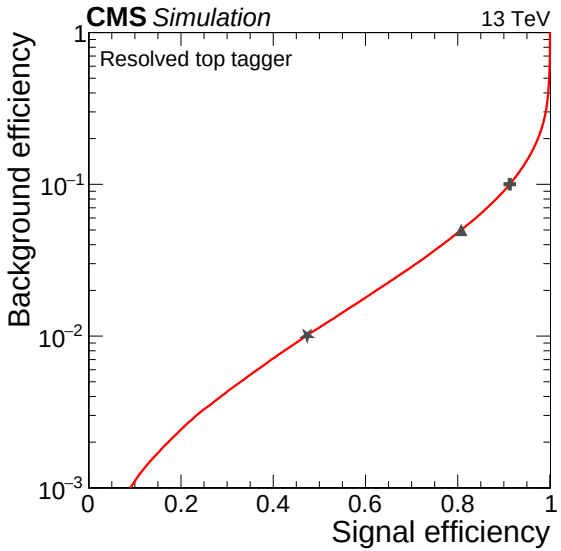
background

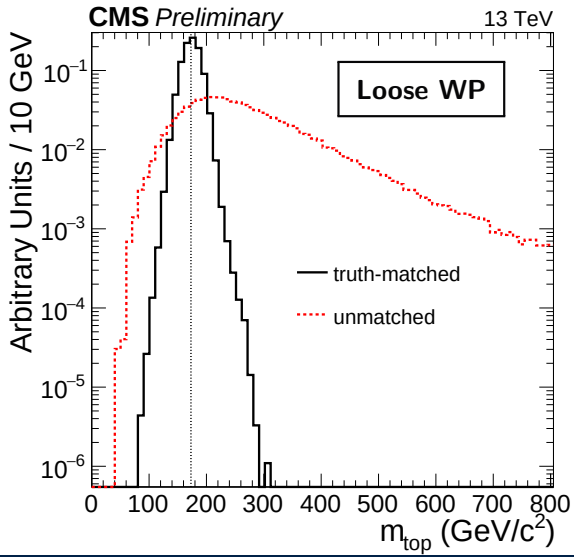


Mass decorrelation using sample reweighting:

- 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 $H^{\pm} \rightarrow$ 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

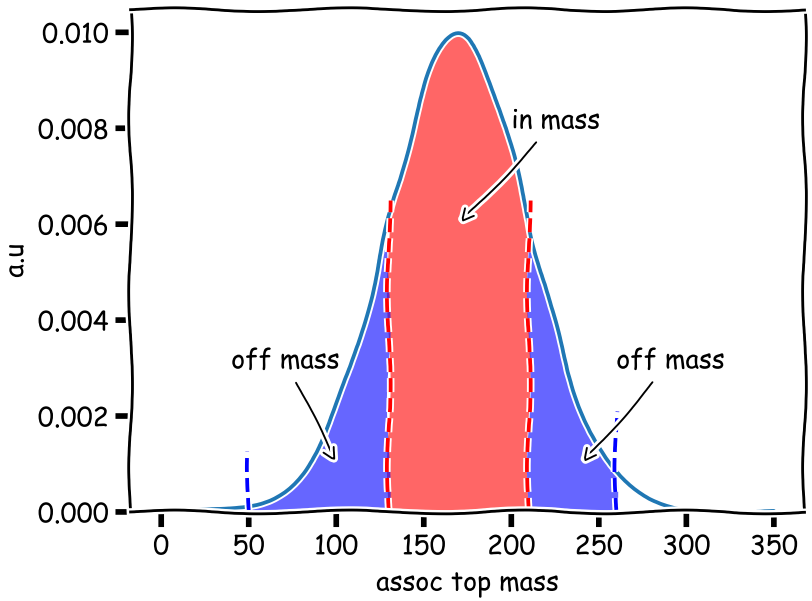
- t^{res}_{assoc} mass: In-mass $\rightarrow$ Off-mass "sidebands"
- $t^{res}_{H^{\pm}}$ mva: t-tagged (t) $\rightarrow$ non t-tagged (!t)

"ABCD" method

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

- Performed in 3 bins of the t^{res}_{assoc} p_{T}:
  - $p_{\mathrm{T}} \in [0, 100, 300, \infty]$ GeV — FIXME

Sidebands



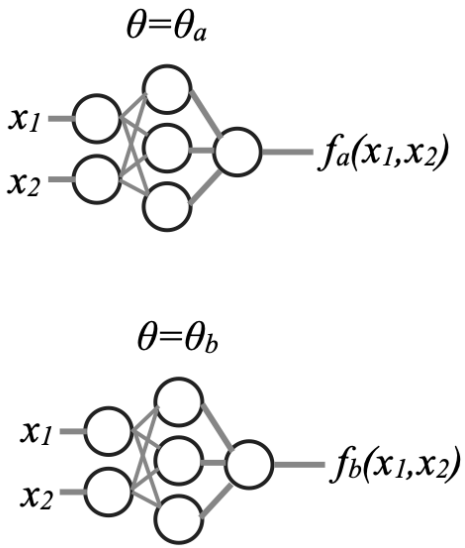
A parameterized DNN is developed to extract signal from SM background

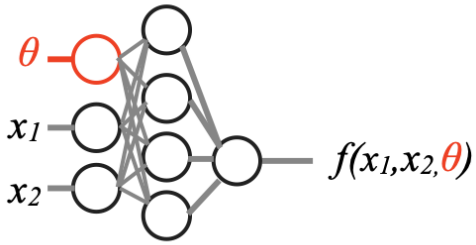
- Signal: $H^{\pm} \rightarrow$ tb for different mass hypotheses
- Background: $t\bar{t} \rightarrow$ SR, Combinatorial $\rightarrow$ CR$^{(off-mass,t)}$ 

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

Signal region (SR): FIXME

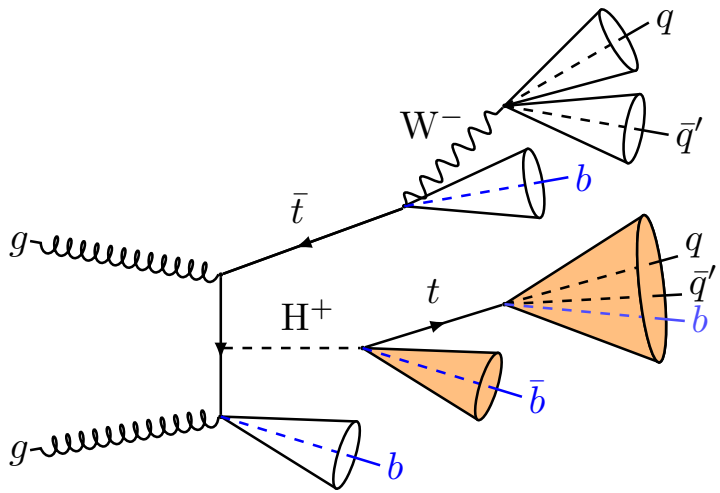
Trigger	H_T + multijet + 1 or 2 b jets H_T + AK8 jet + trim mass
ℓ veto	same as resolved
= 1 boosted top (t^{bst})	loose ID, $p_T > 400$ GeV, $\|\eta\| < 2.4$, $m_{SD} \in [105,210]$ GeV, $\tau_{32} < 0.67$, $N_{b\ \text{subjets}} \geq 1$
$\geq$ 4 jets	$p_T > 40$ GeV, $\|\eta\| < 2.4$, tight ID, $H_T > 500$ GeV
$\geq$ 2 b jets	$p_T > 40$ GeV, DeepJet Medium WP
$\geq 1\ t^{res}$	same as resolved
Kinematic requirements	$\Delta R(t^{bst}, b^{ldg}) > 1.0$ $\max(m_{bb}) > 200$ GeV

SR categorization FIXME

-

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

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



BACKUP