THE HENRYK NIEWODNICZAŃSKI
INSTITUTE OF NUCLEAR PHYSICS
POLISH ACADEMY OF SCIENCES

How to look for a charged Higgs in ATLAS data. The MVA approach

Kraków, 11.01.24

EPIPHANY Conference

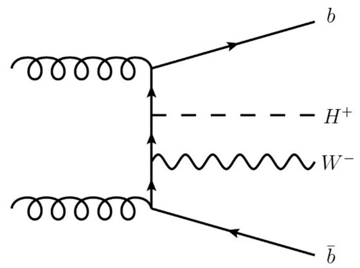
Zuzanna Żak on behalf of
the ATLAS collaboration

Motivation

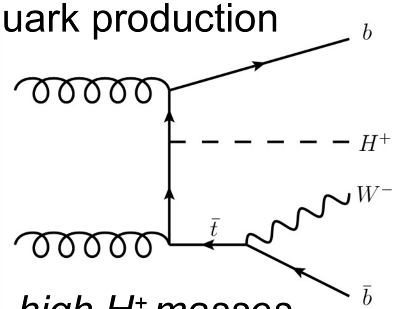
- Is the Higgs boson discovered at LHC the one predicted by the SM, or a part of an extended scalar sector?
- In the Two-Higgs-Doublet-Model, 2HDM (an extension of the SM), there are predicted five Higgs states:
  - Two charged bosons $H^{\pm}$
  - Two CP-even neutral bosons h (observed at the LHC) and H
  - One CP-odd neutral boson A
- **A discovery of the H^{+} would be a clear evidence of physics beyond the Standard Model.**

H^+ production

non-resonant top quark production

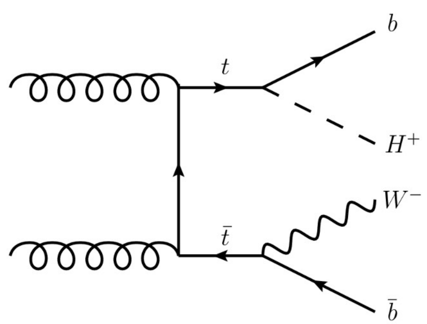


single-resonant top quark production

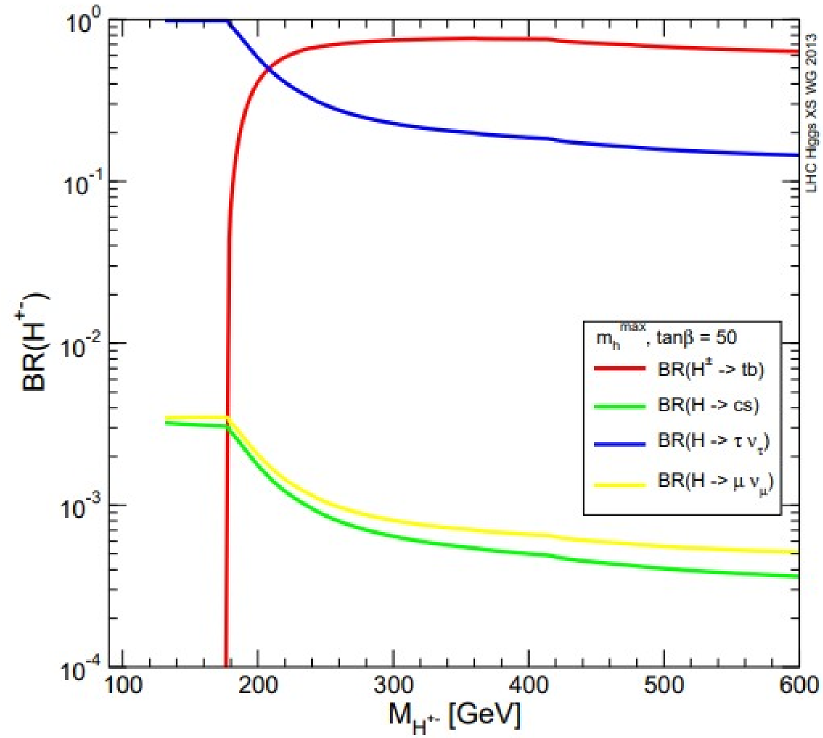


high H^+ masses

double-resonant top quark production



low H^+ masses



Branching ratios of the H^+ decays as a function of m_{H+} for tanβ=50
arXiv:1307.1347

Search for charged Higgs bosons decaying via $H^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at √s=13 TeV with the ATLAS experiment

JHEP 09 (2018) 139

- Signal mass range: 90-2000 GeV.
- Two channels are targeted, depending on the decay mode of the associated W boson: **τ+jets** and **τ+leptons.**

τ + jets

Sensitive at large m_{H+}.

Event selection:

- Electron and muon veto,
- E_{tmiss} >150 GeV,

τ + lep

Sensitive at low and intermediate m_{H+}.

Event selection:

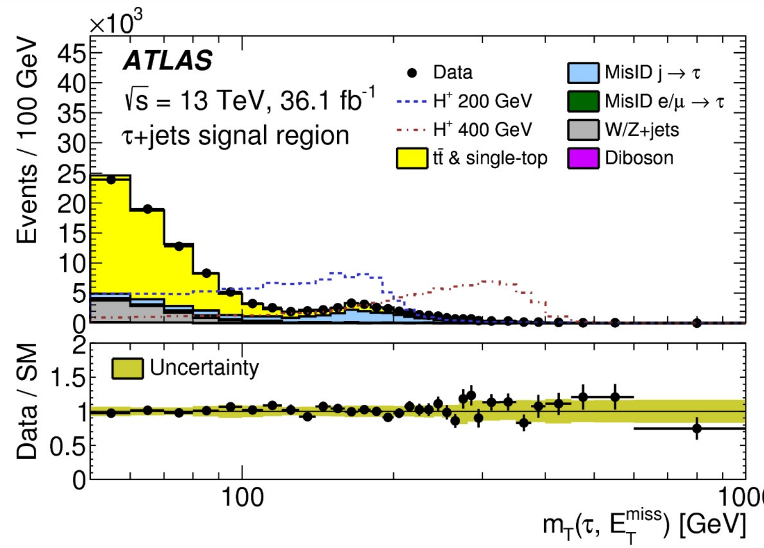
- 1 lepton with p_T>30 GeV
- E_{tmiss} >50 GeV,
- Lepton and τ with opposite signs.

JHEP 09 (2018) 139

Backgrounds JHEP 09 (2018) 139

Sample	Event yields $\tau_{\text{had-vis}}$+jets		
True τ_{had}			
$t\bar{t}$	6900	± 60	± 1800
Single-top-quark	750	± 20	± 100
$W \to \tau\nu$	1050	± 30	± 180
$Z \to \tau\tau$	84	± 42	± 28
Diboson (WW, WZ, ZZ)	63.2	± 4.6	± 7.2
Misidentified $e, \mu \to \tau_{\text{had-vis}}$	265	± 12	± 35
Misidentified jet $\to \tau_{\text{had-vis}}$	2370	± 20	± 260
All backgrounds	11500	± 80	± 1800
H^+ (170 GeV), hMSSM $\tan\beta = 40$	1400	± 10	± 170
H^+ (1000 GeV), hMSSM $\tan\beta = 40$	10.33	± 0.06	± 0.78
Data	11021		

- Expected event yields for the backgrounds and a hypothetical H^+ signal.



- Distribution of m_T for the background and signal corresponding to the H^+ mass 200 and 400 GeV.

Dominant backgrounds: ttbar production, jets misidentified as fake τ.
Backgound estimation:

- Backgrounds with a true τ: MC
- Backgrounds with leptons faking τ: MC +data driven corrections
- Backgrounds with jets faking τ: data-driven fake-factor method

Since the H^+ signal is weak compared to the SM backgorunds, an effective way of signal-background separation is needed.

Multivariate Analysis approach JHEP 09 (2018) 139

- The output score of the machine learning algorithm is used to separate the signal from the SM background processes.
- Machine learning algorithm used: Boosted Decision Trees (BDTs)
- Separate trainings for τ+jets and τ+lep channels
- Training in 5 H^+ mass bins.
- Polarisation variable ϒ is used for 1-prong τ objects with m_{H+} <500 GeV

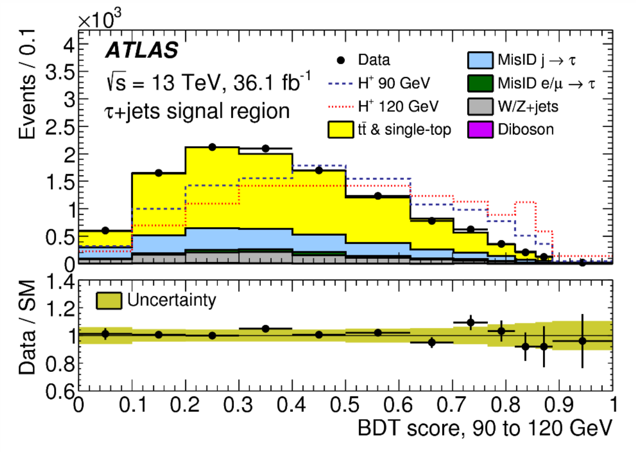
BDT input variable	$\tau_{\text{had-vis}}$+jets	$\tau_{\text{had-vis}}$+lepton
E_T^{miss}	✓	✓
p_T^{τ}	✓	✓
$p_T^{b\text{-jet}}$	✓	✓
p_T^{ℓ}		✓
$\Delta\phi_{\tau,\text{ miss}}$	✓	✓
$\Delta\phi_{b\text{-jet, miss}}$	✓	✓
$\Delta\phi_{\ell,\text{ miss}}$		✓
$\Delta R_{\tau,\ell}$		✓
$\Delta R_{b\text{-jet},\ell}$		✓
$\Delta R_{b\text{-jet},\tau}$	✓	
Υ	✓	✓

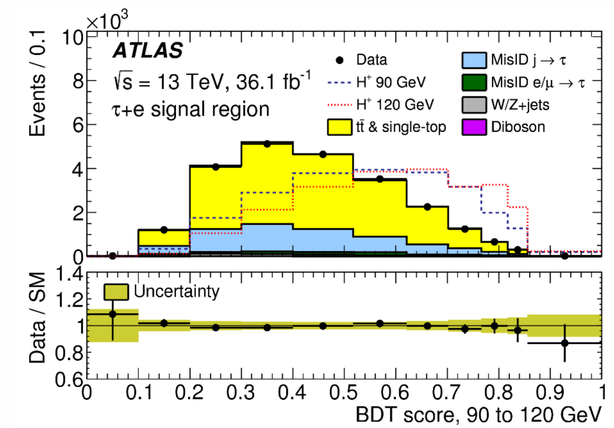
$$\Upsilon = \frac{E_T^{\pi^\pm} - E_T^{\pi^0}}{E_T^{\tau}} \approx 2\frac{p_T^{\tau\text{-track}}}{p_T^{\tau}} - 1$$

Boosted Decision Trees scores JHEP 09 (2018) 139

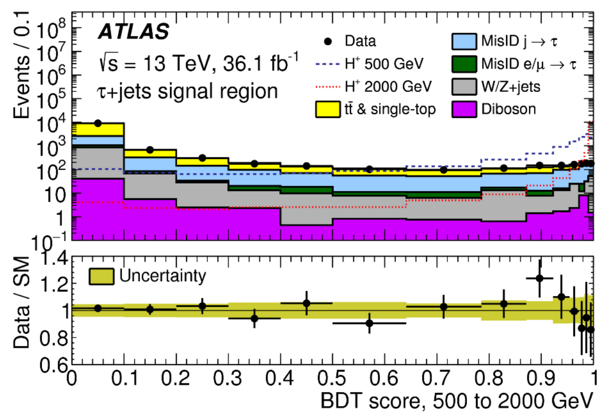
τ+jets signal region

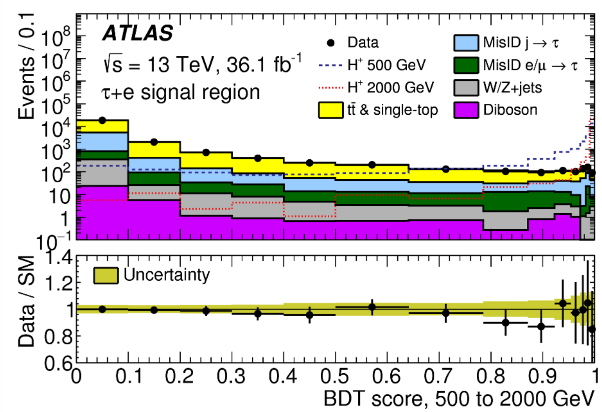
τ+e signal region





H^+ mass: 90-120 GeV

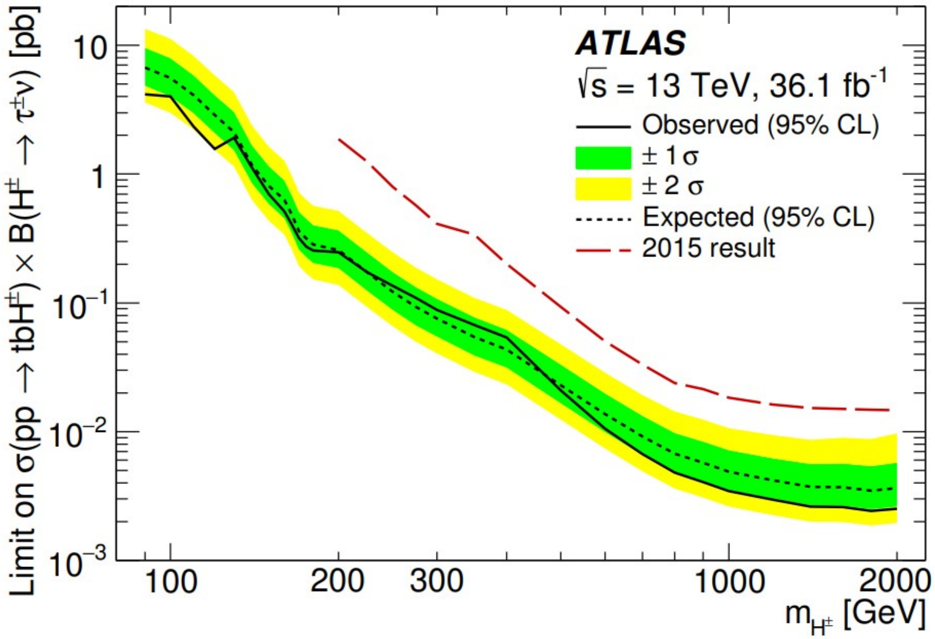


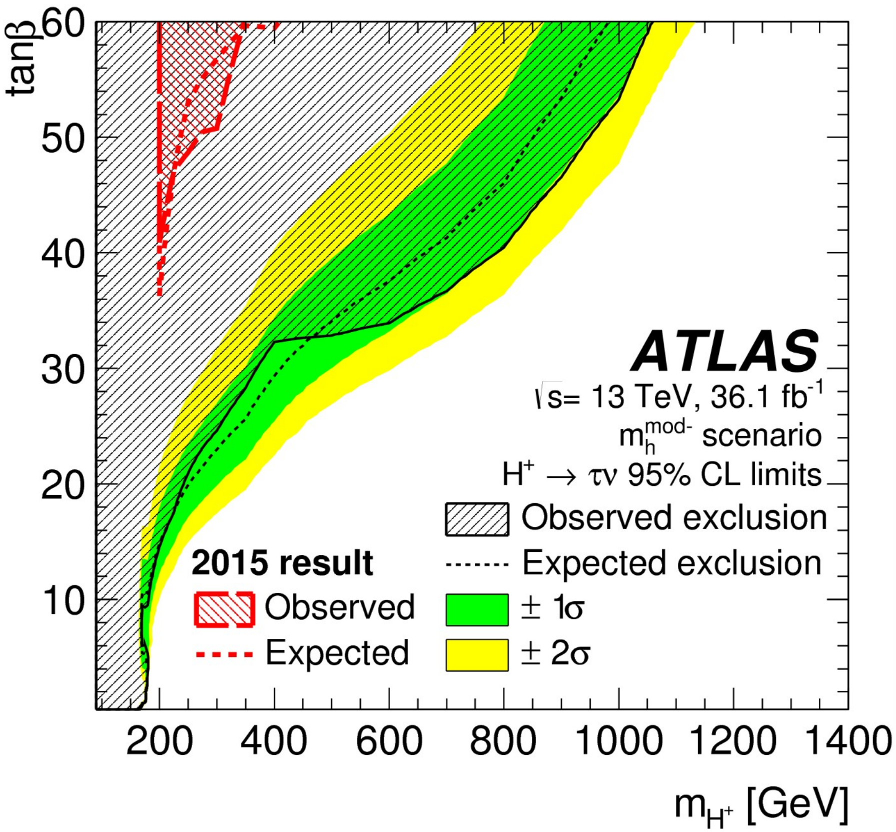


H^+ mass: 500-
- 2000 GeV

$H^+ \rightarrow \tau\nu$ analysis – results JHEP 09 (2018) 139

Systematic uncertainties:
dominant at low m_{H+}: fake factors method
dominant at high m_{H+} : signal modelling

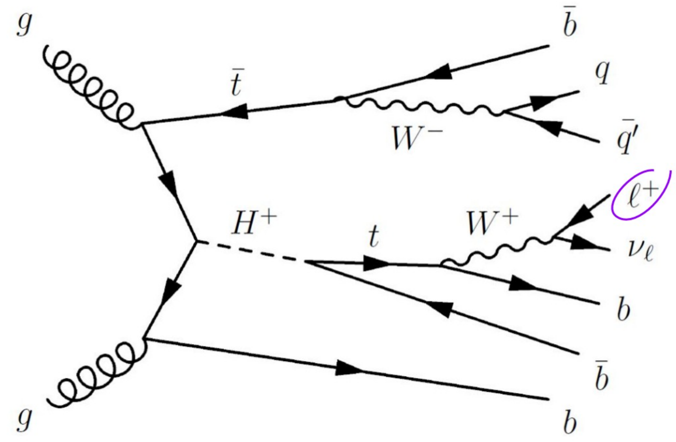




No statistically significant deviation from the SM predictions found.

Search for charged Higgs bosons decaying into a top quark and a bottom quark at √s = 13 TeV with the ATLAS detector JHEP 06 (2021) 145

- 139 fb^{-1} of data from the whole Run2.
- H^{+} mass range 200-2000 GeV.
- Single lepton channel targeted.
- **Event selection:**
  - Exactly one lepton (e^{+} or μ^{+}).
  - $\geq$ 5 jets, $\geq$ 2 b-tagged at 70% efficiency.
- **Classify events according to jet and b-jet multiplicities:**
  - Four signal regions: 5j3b, 5j≥4b, ≥6j3b, ≥6j≥4b.

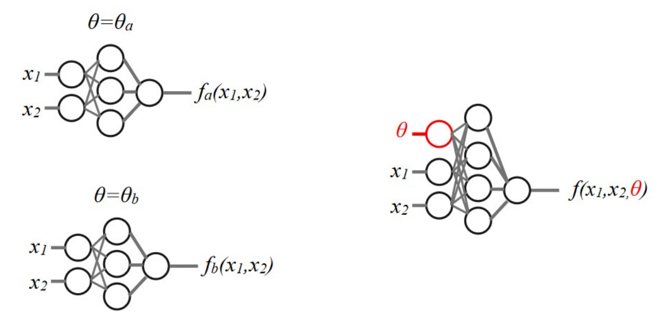


Dominant backgrounds: ttbar + jets
Similarly to the H^{+}→ τν process, the high amount of the backgrounds requires an effective method of signal-background separation.

JHEP 06 (2021) 145

Parametrised Neural Networks [JHEP 06 (2021) 145]

- This kind of algorithm is able to smoothly interpolate between the values of the parameter (in this case, the mass of H^+).
- High-level kinematic variables (e.g. leading jet p_T) are used as an input to the pNN algorithm.
- A single training is performed for each signal region.
- All H^+ samples are included in a single training:
  - Simplifies training, benefits from the continuity, effectively more signal statistics, allows interpolation.

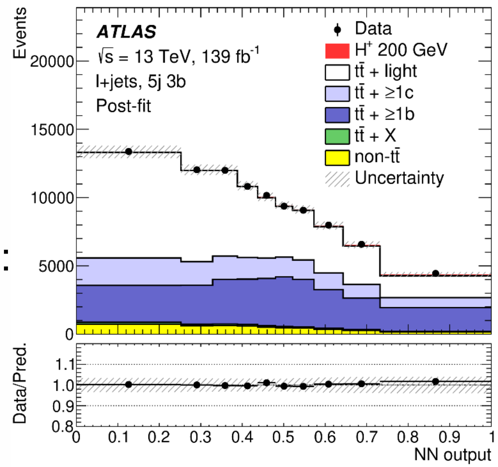


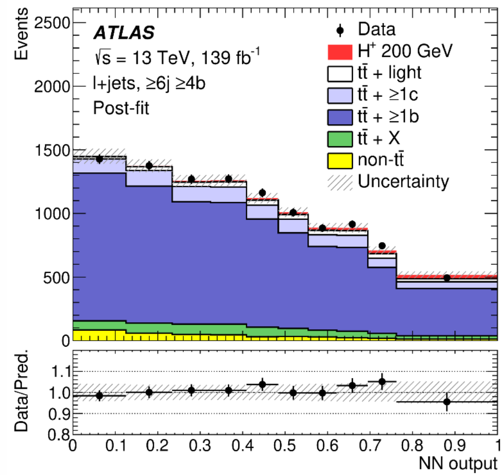
Instead of training the network for each value of the parameter „θ", this parameter is added to the list of input variables.
https://doi.org/10.1140/epjc/s10052-016-4099-4

PNN scores [JHEP 06 (2021) 145]

H⁺ mass: 200 GeV

Signal region: 5j3b



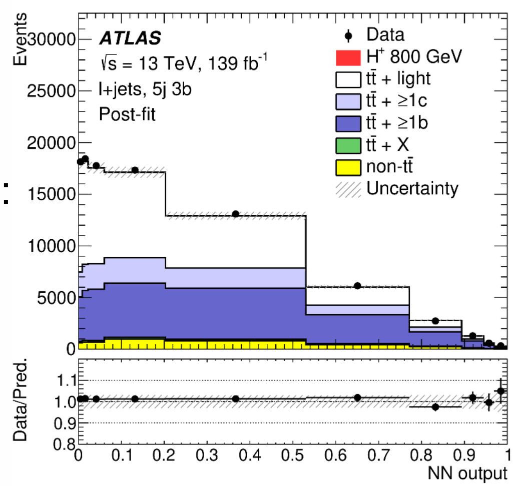


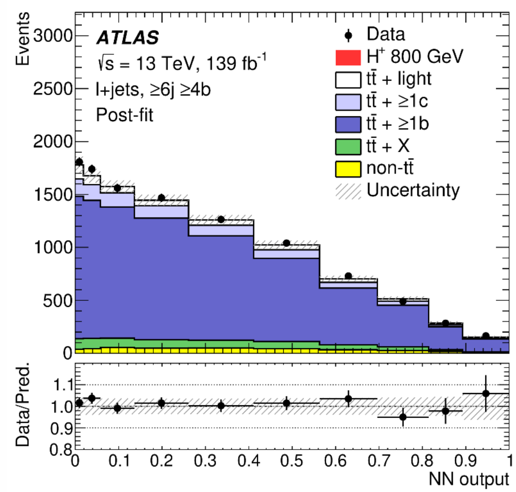
H⁺ mass: 200 GeV

Signal region: >6j≥4b

H⁺ mass: 800 GeV

Signal region: 5j3b



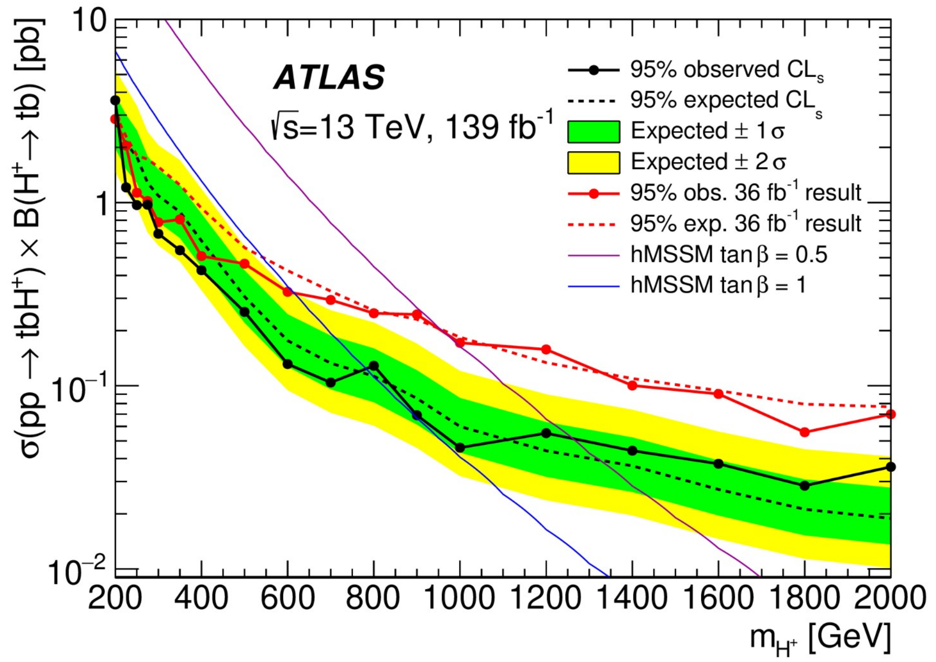


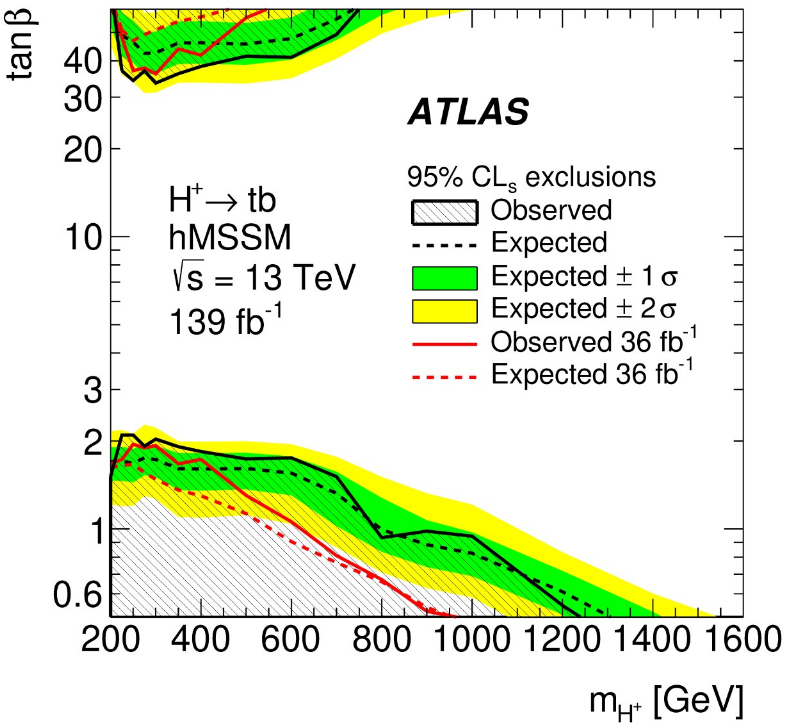
H⁺ mass: 800 GeV

Signal region: >6j≥4b

$H^+ \rightarrow tb$ analysis – results

[JHEP 06 (2021) 145]

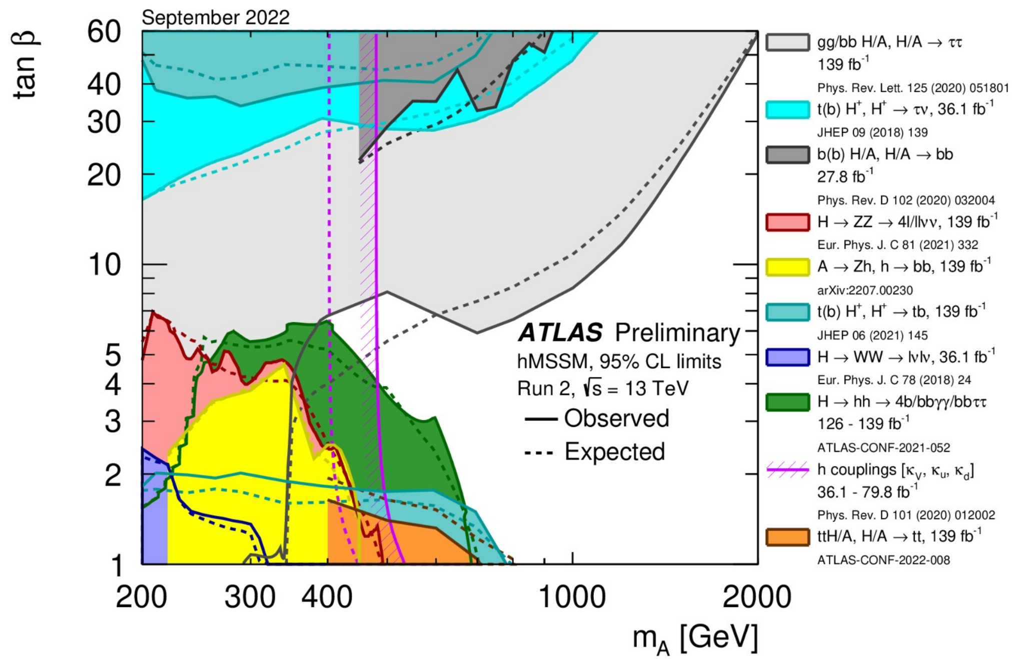




- No statistically significant deviation from the SM predictions found.
- Dominant systematic uncertainties: ttbar≥1b modelling.

Summary

- Finding the H^+ would be an evidence for the physics beyond the Stanard Model.
- The searches for H^+ are targeting different production and decay modes, such as $H^+ \to \tau\nu$, $H^+ \to tb$.
- Both analyses helped to update the exclusion limits on the corresponding decays.
- MVA proves to be a valuable tool for data analysis in this type of searches.

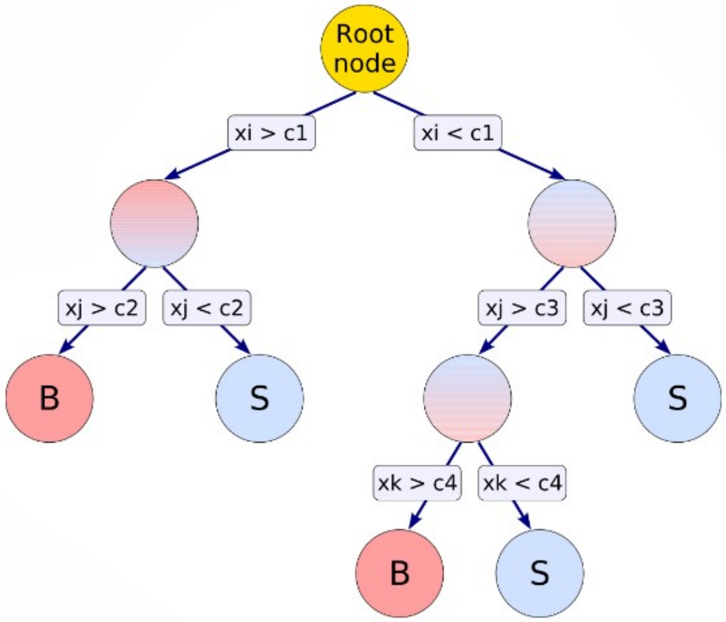


Exclusion limits on tan β in the context of hMSSM. Results from both described analyses are included. ATL-PHYS-PUB-2022-043

BACKUP SLIDES

Boosted Decision Trees

- The goal of the agorithm is to predict whether the event belongs to the „signal” or „background” class.
- After each of the tree nodes, a cut is applied which produces two daughter nodes.
- After the stopping condition is reached, a purity value is assigned to the final node.



$H^+ \rightarrow tb$ background composition

m_{H^+} = 200 GeV hypothesis				
	5j, 3b	5j, ≥ 4b	≥ 6j, 3b	≥ 6j, ≥ 4b
$t\bar{t}$ + light	45000 ± 4000	310 ± 110	32000 ± 4000	340 ± 140
$t\bar{t}$ + ≥1b	29600 ± 2900	2940 ± 220	40200 ± 3300	8000 ± 500
$t\bar{t}$ + ≥1c	14000 ± 4000	440 ± 140	19000 ± 6000	1010 ± 290
$t\bar{t}$ + W	110 ± 15	3.2 ± 0.6	236 ± 35	16.2 ± 2.7
$t\bar{t}$ + Z	300 ± 40	51 ± 6	670 ± 90	174 ± 23
Single-top Wt-channel	2300 ± 600	80 ± 50	1900 ± 800	150 ± 90
Single-top t-channel	740 ± 300	51 ± 20	500 ± 400	60 ± 50
Other top-quark sources	128 ± 16	17.5 ± 3.2	180 ± 70	58 ± 24
VV & V + jets	1600 ± 600	65 ± 23	1600 ± 600	120 ± 40
$t\bar{t}H$	530 ± 60	127 ± 19	1140 ± 120	430 ± 60
H^+	600 ± 900	70 ± 90	700 ± 1000	160 ± 230
Total	95700 ± 2900	4150 ± 140	98400 ± 2900	10500 ± 400
Data	95852	4109	98929	10552

m_{H^+} = 800 GeV hypothesis				
	5j, 3b	5j, ≥ 4b	≥ 6j, 3b	≥ 6j, ≥ 4b
$t\bar{t}$ + light	46000 ± 4000	330 ± 120	33000 ± 4000	500 ± 200
$t\bar{t}$ + ≥1b	29600 ± 3100	2920 ± 210	41000 ± 4000	8100 ± 400
$t\bar{t}$ + ≥1c	14000 ± 6000	440 ± 190	17000 ± 7000	870 ± 330
$t\bar{t}$ + W	108 ± 15	3.3 ± 0.6	233 ± 35	16.0 ± 2.7
$t\bar{t}$ + Z	300 ± 40	50 ± 7	660 ± 90	171 ± 23
Single-top Wt-channel	2000 ± 500	56 ± 33	1400 ± 500	100 ± 60
Single-top t-channel	740 ± 300	53 ± 21	600 ± 500	70 ± 50
Other top-quark sources	130 ± 16	17.7 ± 3.2	190 ± 70	61 ± 24
VV & V + jets	1900 ± 700	73 ± 25	1700 ± 600	130 ± 50
$t\bar{t}H$	520 ± 60	125 ± 19	1130 ± 120	420 ± 60
H^+	30 ± 80	4 ± 10	70 ± 180	20 ± 50
Total	94700 ± 2800	4070 ± 140	97800 ± 2800	10400 ± 400
Data	95852	4109	98929	10552

$H^+ \to \tau\nu$ analysis – kinematic variables used for the MVA

1. E_T^{miss} - missing transverse energy;
2. p_T^τ - visible transverse momentum of τ lepton; in case of more τ_{had} candidates the one with the highest p_T^τ is chosen;
3. p_T^{b-jet} - transverse momentum of b-tagged jet; in case of more b-jets the one with the highest p_T^{b-jet} was chosen;
4. $\Delta\Phi_{\tau_{had-vis},E_T^{miss}}$ - difference in azimuthal angle between a reconstructed hadronically decaying τ lepton and the direction of the missing transverse momentum;
5. $\Delta\Phi_{b-jet,E_T^{miss}}$ - difference in azimuthal angle between a reconstructed b-tagged jet and the direction of the missing transverse momentum;
6. $\Delta R_{b-jet,\tau_{had-vis}}$ - difference of radius parameter between a reconstructed b-tagged jet and a reconstructed hadronically decaying τ lepton.

$$\Upsilon = \frac{E_T^{\pi^\pm} - E_T^{\pi^0}}{E_T^\tau} \approx 2\frac{p_T^{\tau-track}}{p_T^\tau} - 1,$$

where: $E_T^{\pi^{\pm,0}}$ - energies carried by the charged and neutral pions in the 1-prong τ lepton decay, $p_T^{\tau-track}$ - transverse momentum of the track associated with τ_{had}, E_T^τ - transverse energy of the τ lepton. This variable can be only defined for the events where τ_{had} has only one associated track.

H^+ →tb analysis – kinematic variables used for the MVA

- kinematic discriminant D defined in the text,
- scalar sum of the p_T of all jets,
- centrality calculated using all jets and leptons,
- p_T of the leading jet,
- p_T of fifth leading jet,
- invariant mass of the b-jet pair with minimum ΔR,
- invariant mass of the b-jet pair with maximum p_T,
- largest invariant mass of a b-jet pair,
- invariant mass of the jet triplet with maximum p_T,
- invariant mass of the untagged jet-pair with minimum ΔR (not in 5j≥4b),
- average ΔR between all b-jet pairs in the event,
- ΔR between the lepton and the pair of b-jets with smallest ΔR,
- second Fox-Wolfram moment calculated using all jets and leptons
- number of jets (only in ≥6j3b and ≥6j≥4b regions), and
- number of b-jets (only in 5j≥4b and ≥6j≥4b regions).

BDT architecture in the $H^+ \rightarrow \tau\nu$ analysis

Hyperparameter	Mass ranges	Mass points
n estimators	100	80
learning rate	0.1	0.1
max depth	10	8
min samples leaf	0.005	0.005
min samples split	0.01	0.02

pNN architecture in the $H^+ \rightarrow tb$ analysis

- 2 fully connected layers of 64 nodes
- Activation function: rectified linear unit
- Loss funtion: binary cross-entropy
- Dropout: 10%

Expected distibutions of the pNN output in the $H^+ \rightarrow tb$ analysis