THE HENRYK NIEWODNICZAŃSKI
INSTITUTE OF NUCLEAR PHYSICS
POLISH ACADEMY OF SCIENCES

Data-driven estimation of fake τ-lepton background in the $H^{\pm} \rightarrow \tau^{\pm}\nu$ search with ATLAS Run-II data

Paper: Phys. Rev. D 111 (2025) 072006

Monika Juzek (INP PAS)

On behalf of the ATLAS Collaboration

Bologna, May 6th 2026

DIS2026

ATLAS EXPERIMENT

Motivation - the need for physics beyond the Standard Model

Discovery of the Higgs boson in 2012 (ATLAS & CMS) confirmed the Standard Model (SM) expectations by providing the last misisng element to the framework.

Still SM has some limitations

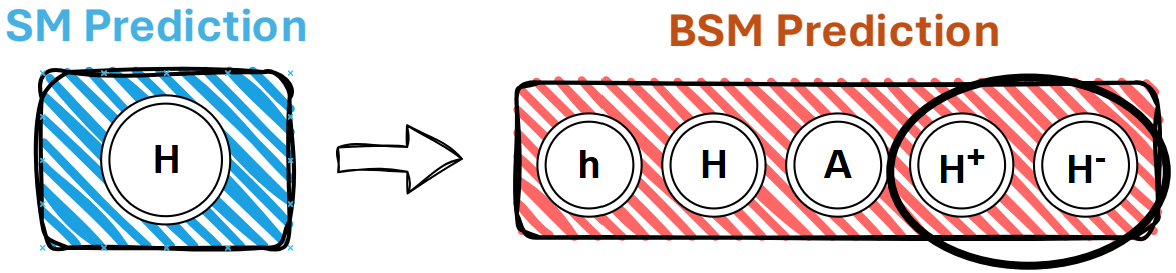
→ neutrino masses, baryon asymmetry, dark matter & dark energy

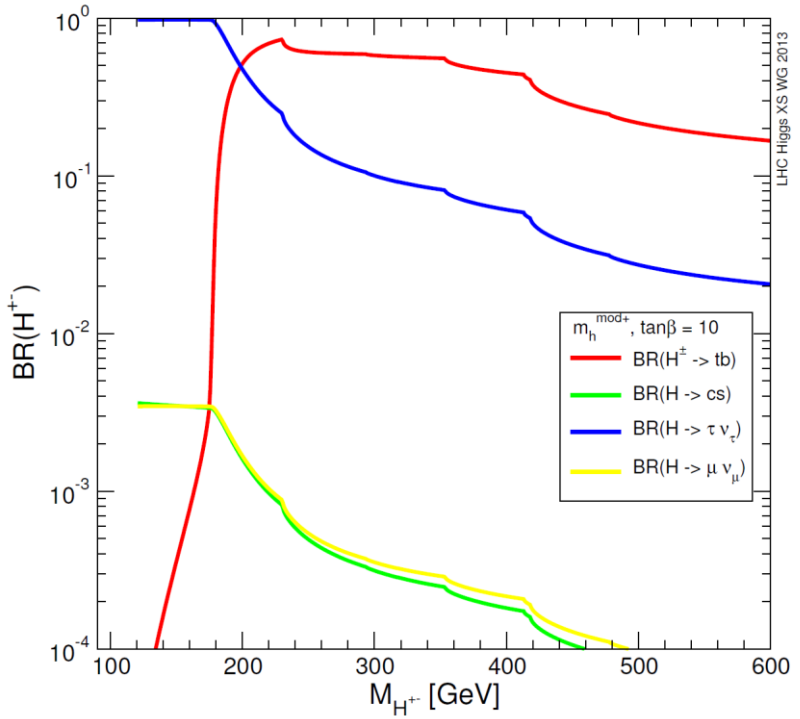
New Physics extensions of the scalar sector, for instance:

- **2HDM (Two-Higgs-Doublet Models)**
- **MSSM (Minimal Supersymmetric Standard Model)**

Prediction of new states

→ additional Higgs doublets with new bosons, including **charged Higgs (H±)**





[1]

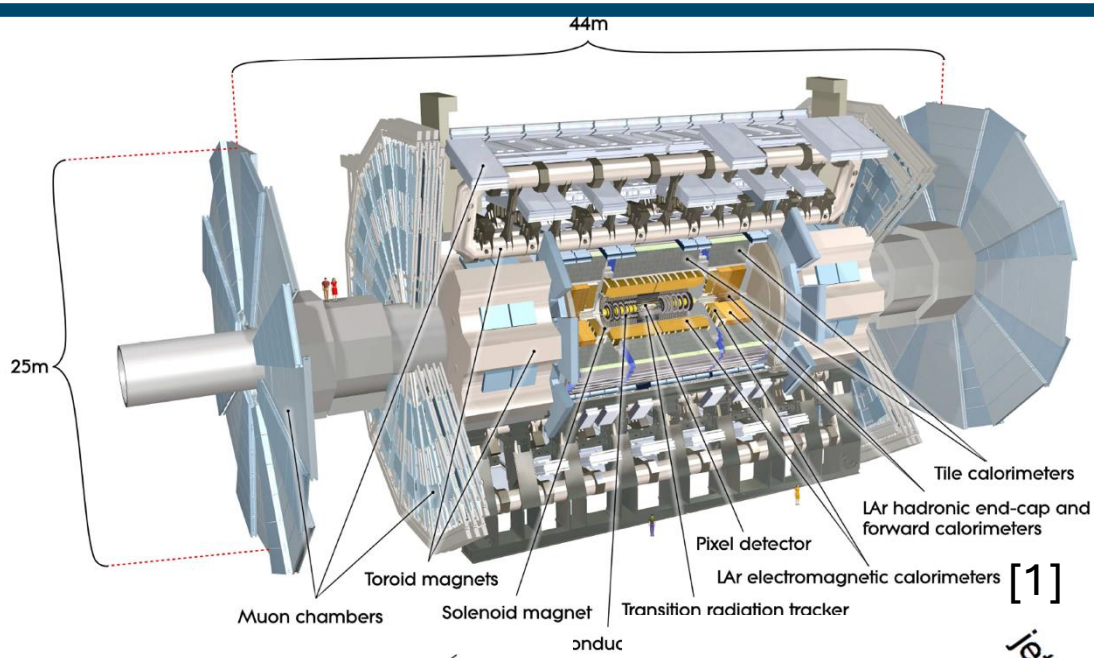
[1] S. Heinemeyer et al., Handbook of LHC Higgs Cross Sections: 3. Higgs Properties, (2013), arXiv: 1307.1347.

The τ lepton in the ATLAS experiment

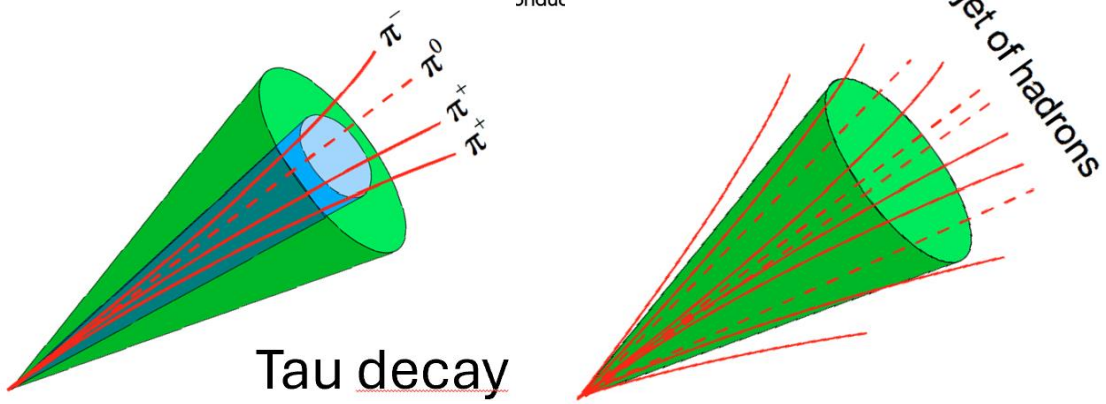
Searches for new BSM phenomena are conducted at the **Large Hadron Collider (LHC)**

ATLAS is the largest general-purpose detector

- **τ lepton - only lepton with hadronic decays**
  → Decay channels: ~65% hadronic, ~35% leptonic
- **ATLAS identification is challenging due to jet-like signature**
- **Techniques -** machine learning (RNN for Run-II) using tracker & calorimeter info
- Plays a key role in Higgs studies and searches for new physics



[1]



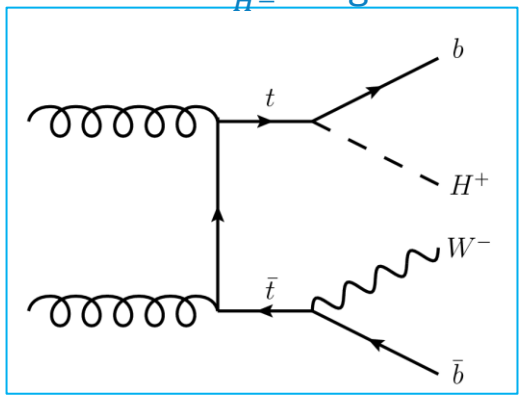
[1] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. JINST 3 (2008), S08003.

$H^{\pm} \rightarrow \tau^{\pm}\nu$ analysis using Run-II data

- **Search for charged $H^{\pm}$ decaying to τν using 140 fb^{-1} of $\sqrt{s}$ = 13 TeV pp data collected by the ATLAS experiment at the LHC during the 2015-2018**
- $H^{\pm}$ production process depends on the mass range but is typically associated with t and b quarks
- **Investigated signal mass range: 80 GeV ≤ $m_{H^{\pm}}$ ≤ 3000 GeV**

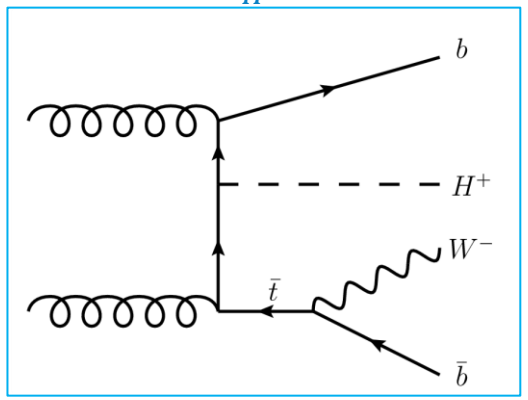
  → precise calibration of the Tau Energy Scale becomes a critical challenge at high $m_{H^{\pm}}$
- Analysis split into two sub-channels based on the decay mode of the associated t:
  - **τ+jets:** most sensitive to high $H^{\pm}$ mass
  - **τ+lepton:** most sensitive to low and intermediate $H^{\pm}$ mass
- **MVA:** Parametrized Neural Network (PNN) score used as the final discriminant in a likelihood fit

Low $m_{H^{\pm}}$ range



[1]

High $m_{H^{\pm}}$ range



[1]

Backgrounds with true τ:

- Modelled using MC ($t\bar{t}$, W+jets, Z+jets and VV+jets)

Background with fake τ candidates:

- $lepton \rightarrow \tau$: estimated using MC, validated on $Z \rightarrow ee$
- jets → ***τ: Fake Factor method (FF)***

[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)

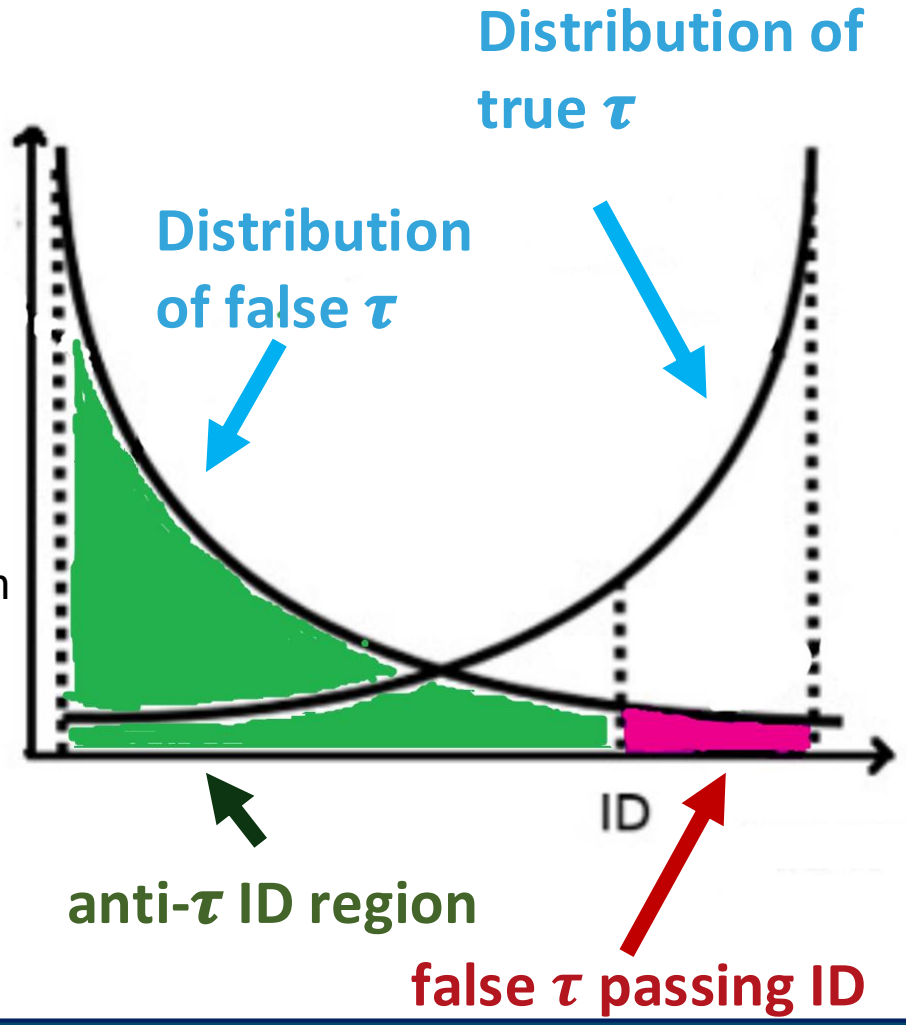
Fake Factor method

- Data-driven method
- For the control region (CR) dominated by the fake-τ background we define **fake factors (FF) as:**

$$FF = \frac{N^{CR}_{\tau-ID}}{N^{CR}_{anti_\tau-ID}}$$

$\rightarrow$ Jets that pass τ nominal object selection (**τ ID**)

$\rightarrow$ Jets that fail the nominal τ identification (**anti-τ ID**)



- **Total number of background events from jets** (N^{τ}_{fakes}) **in signal region:**

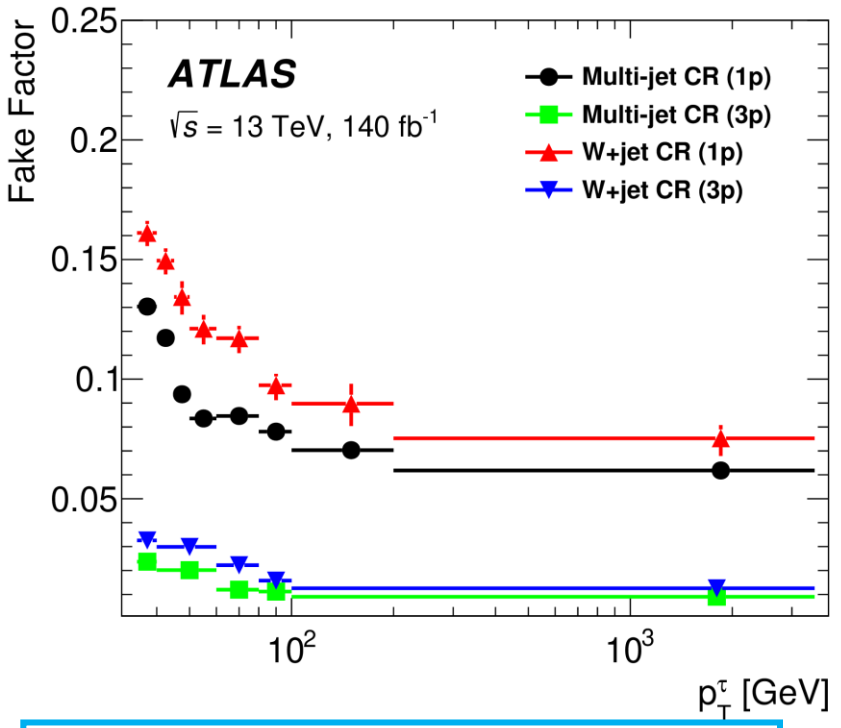
$$N^{\tau}_{fakes} = FF \times N^{SR}_{anti_\tau-ID}$$

Fake Factor method in $H^{\pm} \rightarrow \tau^{\pm}\nu$ analysis

- **FFs are extracted in the two CRs (selection on slide 12):**
  - I. **Multi-jets CR** → similar fractions of quark- and gluon-initiated jets
  - II. **W+jets CR** → enriched in the quark-initiated jets
- Parametrization of the FFs : p_T^{τ} and number of charged tracks (1-track or 3-track τ)
- Subtraction of true tau from SM proccesses using MC
- To obtain proper jets composition in a signal region, FFs from two CRs are combined using the formula:

$$\mathrm{FF}_{\mathrm{SR}}^{\mathrm{comb}}(i) = \alpha_{\mathrm{MJ}}(i) \times \mathrm{FF}_{\mathrm{MJ}}(i) + (1 - \alpha_{\mathrm{MJ}}(i)) \times \mathrm{FF}_{\mathrm{W+jets}}(i)$$

α_{MJ} parameter is calculated using a template-fit method



[1]

Tau jet width:

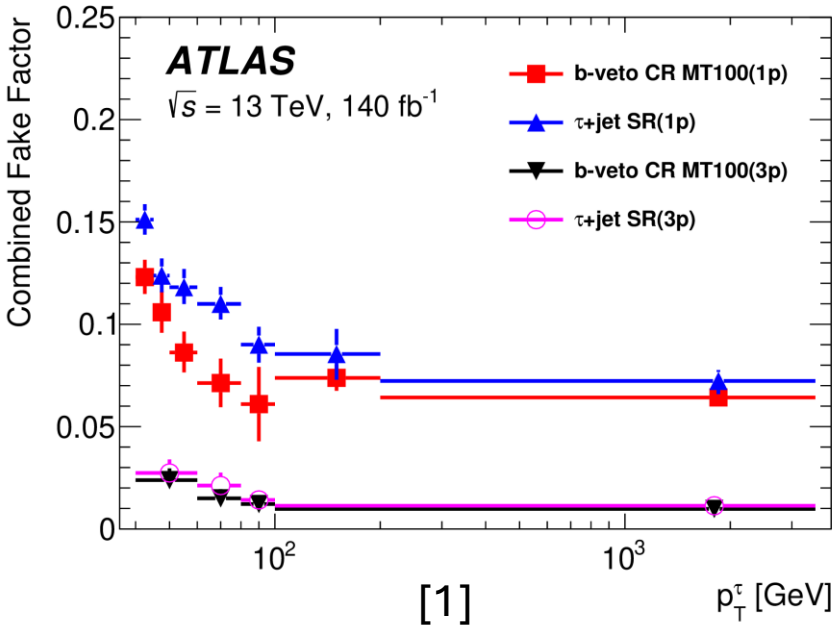
$$w_{\tau} = \frac{\sum \left[p_{\mathrm{T}}^{\mathrm{track}} \times \Delta R\left(\tau_{\mathrm{had-vis}}, \mathrm{track}\right) \right]}{\sum p_{\mathrm{T}}^{\mathrm{track}}}$$

$$\Delta R\left(\tau_{\mathrm{had-vis}}, \mathrm{track}\right) < 0.4.$$

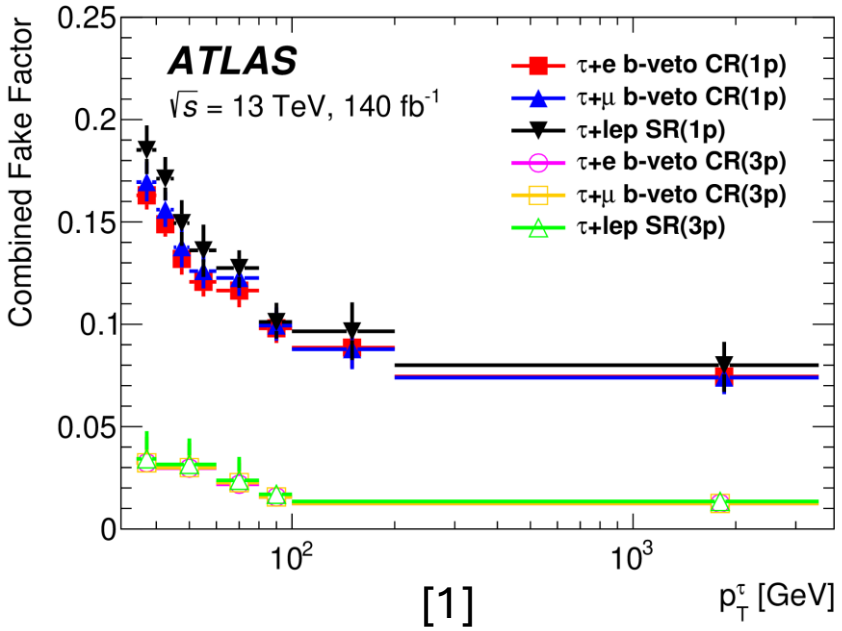
[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)

Obtained FF in the signal regions

τ+jets channel



τ+lepton channel



- Validation of the FF method is done in the regions enriched in the fake τ background (**with b-veto VRs**)

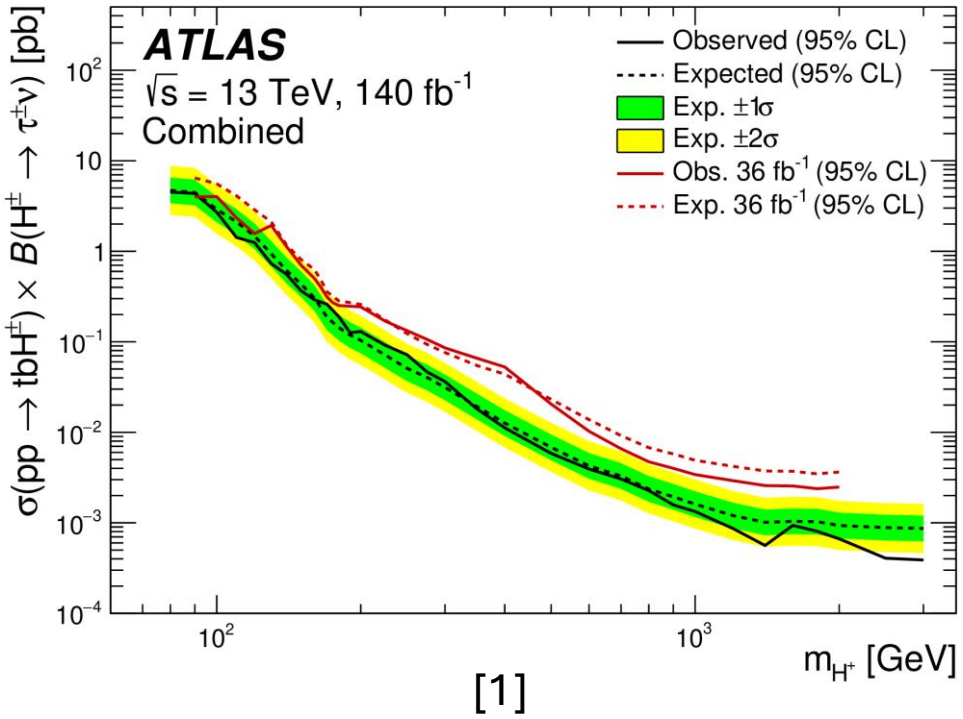
The largest source of systematic uncertainty in the FF method was found to be the contribution from jets originating from heavy (b and c) quarks (≈ 6% in both channels)

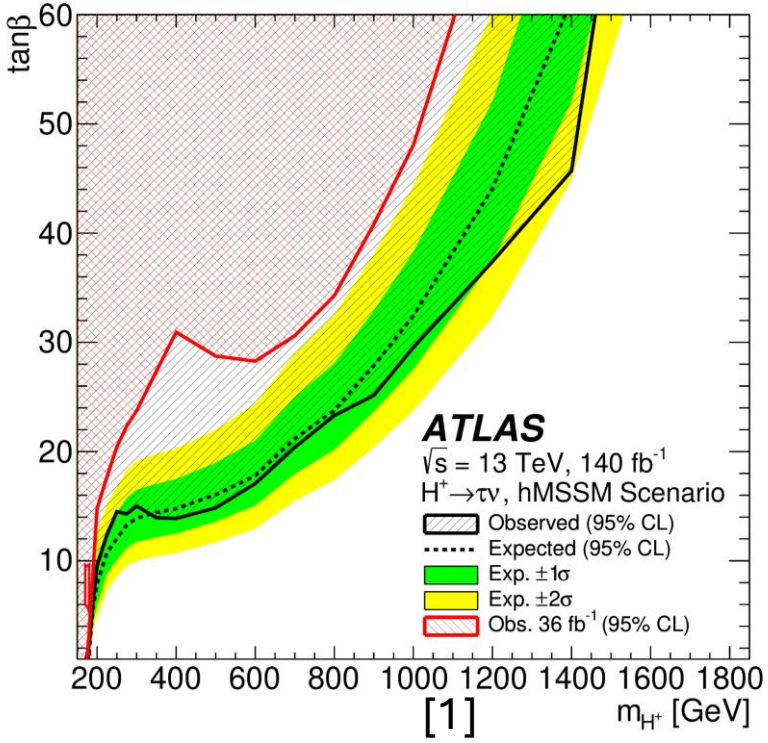
- The estimation of the background from jets misidentified as τ leptons is a significant source of systematic uncertainty across the full $H^{\pm}$ mass range

[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)

Results

- PNN score (with $H^{\pm}$ mass as an input) used as a final discriminant in a likelihood fit





- The plots compare the exclusions obtained in the current analysis (black lines) and the previous analysis (red lines).

[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)

Summary

- **τ lepton** plays a key role in BSM searches at ATLAS
  → powerful probe, but **challenging to reconstruct and identify**

- **Search for $H^+ \rightarrow \tau\nu$ using Run-II data**
  - **Implemenation of the Fake factor method for data-driven estimation of jet → τ misidentification:**
    → Reduces reliance on MC simulation and associated uncertainties
    → Improves analysis sensitivity
  - **Parameterized neural networks (PNN)** → strong signal–background separation
  - **No excess observed → limits improved by a factor of 2-4 across a wide mass range**
  - Run-III data analysis has started

Thank you for your attention

Backup

Challenges in τ reconstruction

Neutrinos in decay

→ incomplete kinematics (MET only)

Jet → τ similarity

→ narrow jets mimic hadronically decaying τ leptons

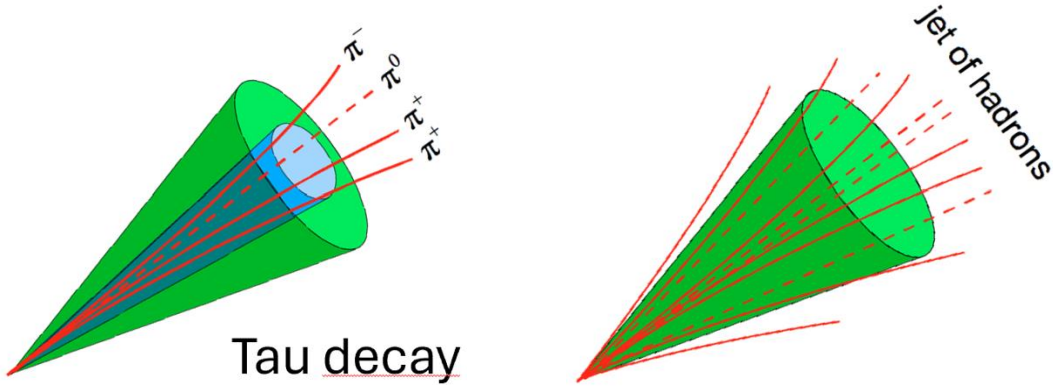
Fake τ background

→ dominant source of uncertainty

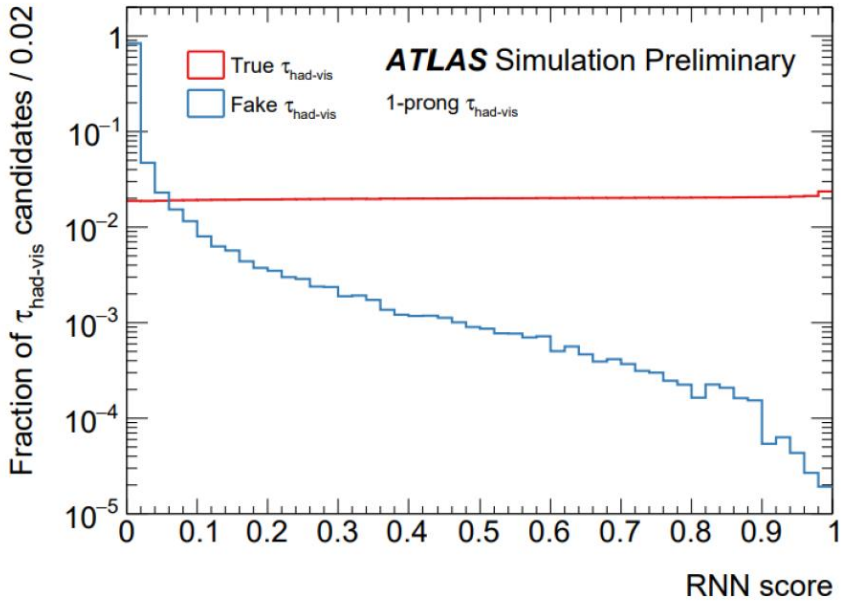
Advanced identification needed

→ ML-based τ ID (for Run 2 data Recurrent Neural Network is used)

Working point	Signal efficiency		Background rejection BDT		Background rejection RNN	
	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong
Tight	60%	45%	40	400	70	700
Medium	75%	60%	20	150	35	240
Loose	85%	75%	12	61	21	90
Very loose	95%	95%	5.3	11.2	9.9	16



Tau decay



[1]

[1] ATLAS Collaboration. Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment. Tech. rep. 2019. url: https://cds.cern.ch/record/2688062/.

Event selection for different regions

Cut	SR $\tau + j$	$t\bar{t}(\tau j)$	$W + j$	FF MJ	FF $W + j$	SR $\tau + \ell$	$t\bar{t}(\tau \ell)$	b-veto	$e + \mu$	VR $\tau + e$ b-veto	VR $\tau + j$ b-veto MT100
$p_{\mathrm{T}}(\tau)$	> 40	> 40	> 40	> 30	> 30	> 30	> 30	> 30	...	> 30	> 40
N_ℓ	0	0	0	0	1	1	1	1	2	1	0
$p_{\mathrm{T}}(\ell)$	...	...	...	...	> 30	> 30	> 30	> 30	> 30	> 30	...
$q(\tau) \times q(\ell)$	...	...	...	...	...	-1	-1	-1	...	-1	...
N_{jet}	≥ 3	≥ 3	≥ 3	≥ 3	...	≥ 1	≥ 2	≥ 1	≥ 1	≥ 1	≥ 3
p_{T}(lead-jet)	> 25	...	> 25	> 25	...	> 25	> 25	> 25	> 25	> 25	> 25
$N_{b\text{-jet}}$	≥ 1	≥ 2	0	0	0	≥ 1	≥ 2	0	≥ 1	0	0
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 150	> 150	> 150	< 80	...	> 50	> 80	> 50	> 50	> 50	> 150
$m_{\mathrm{T}}(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$	> 50	< 100	< 100	> 50	...	...	< 70	...	...	...	> 100
Other	a				b	a			c	a	a

[a] Note that the SRs reject events in the overlap with the corresponding $t\bar{t}$ regions.
[b] There is an additional requirement that 60 GeV $< m_T(e, E_{\mathrm{T}}^{\mathrm{miss}}) <$ 160 GeV.
[c] The two ℓ must be an oppositely charged $e + \mu$ pair.

- **Strict b-jet veto -** Applied in both FF CRs and VRs to ensure orthogonality to the H+ signal and isolate from the $t\bar{t}$ background

- **In addition, kinematic cuts to separate different jet sources:**
  - **Multi-jets FF CR** → Requirement E_T^{miss} < 80 GeV provides gluon-initiated jets
  - **W+jets FF CR** → 60 GeV < $m_T(\ell, E_T^{miss})$ <160 GeV targets quark-initiated jet fractions

Full systematics of Fake Factor method

I. the statistical uncertainties in the event yields entering the computation of FFs (each bin of their parameterization and for each control region)

II. the level of contamination of true τ candidates fulfilling the anti- τ selection (varied by 50%)

III. MC modeling of true tau identification - scale factors error of 5% is implemented

IV. the statistical uncertainty of the best-fit value of α_{MJ}

V. The unceratinty of the α_{MJ} fitting strategy

VI. for the Υ distribution (used in RNN-based τ ID) the uncertainty of the inverse transform sampling method (Smirnov transformation for Υ)

VII. the modelling of heavy-flavour jets mimicking τ candidates

MVA Startegy

- **Mass-parameterized NN (PNN):** final discriminant for signal vs SM background
- **Continuous sensitivity:** parametrization in $m_{H^\pm}$ → one model for all mass hypotheses
- Data-driven $jet \rightarrow \tau$ fakes included for fake taus
- **Architecture & training:** separate PNNs (**τ+jets**, **τ+e.** **τ+μ**), k-fold ($k = 5$) to avoid overtraining
- **τ polarization:** Υ variable for 1-prong decays → scalar (H±) vs vector (W/Z) separation. Applied **Smirnov transform** for robust background modelling

PNN input variable	τ_{had}+jets	τ_{had}+lepton
$p_{\text{T}}^{\tau}, \eta^{\tau}, \phi^{\tau}, E^{\tau}$	✓	✓
$p_{\text{T}}^{\ell}, \eta^{\ell}, \phi^{\ell}, E^{\ell}$		✓
$p_{\text{T}}^{b\text{-jet}}, \eta^{b\text{-jet}}, \phi^{b\text{-jet}}, E^{b\text{-jet}}$	✓	✓
$p_{\text{T}}^{\text{jet-1}}, \eta^{\text{jet-1}}, \phi^{\text{jet-1}}, E^{\text{jet-1}}$	✓	✓
$p_{\text{T}}^{\text{jet-2}}, \eta^{\text{jet-2}}, \phi^{\text{jet-2}}$	✓	
$p_{\text{T}}^{\text{jet-3}}, \eta^{\text{jet-3}}, \phi^{\text{jet-3}}$	✓	
$p_{\text{T}}^{\text{jet-2}}$		✓
$E_{\text{T}}^{\text{miss}}, \phi^{E_{\text{T}}^{\text{miss}}}$	✓	✓
Υ (1-prong $\tau_{\text{had-vis}}$ only)	✓	✓
$m_{\text{truth}}^{H^+}$	✓	✓

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Background
Partition 1	Evaluation	Train	Train	Train	Train	Fold 1
Partition 2	Train	Evaluation	Train	Train	Train	Fold 2
Partition 3	Train	Train	Evaluation	Train	Train	Fold 3
Partition 4	Train	Train	Train	Evaluation	Train	Fold 4
Partition 5	Train	Train	Train	Train	Evaluation	Fold 5

[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^\pm \rightarrow \tau^\pm \nu_\tau$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)

PNN Output - Pre-fit

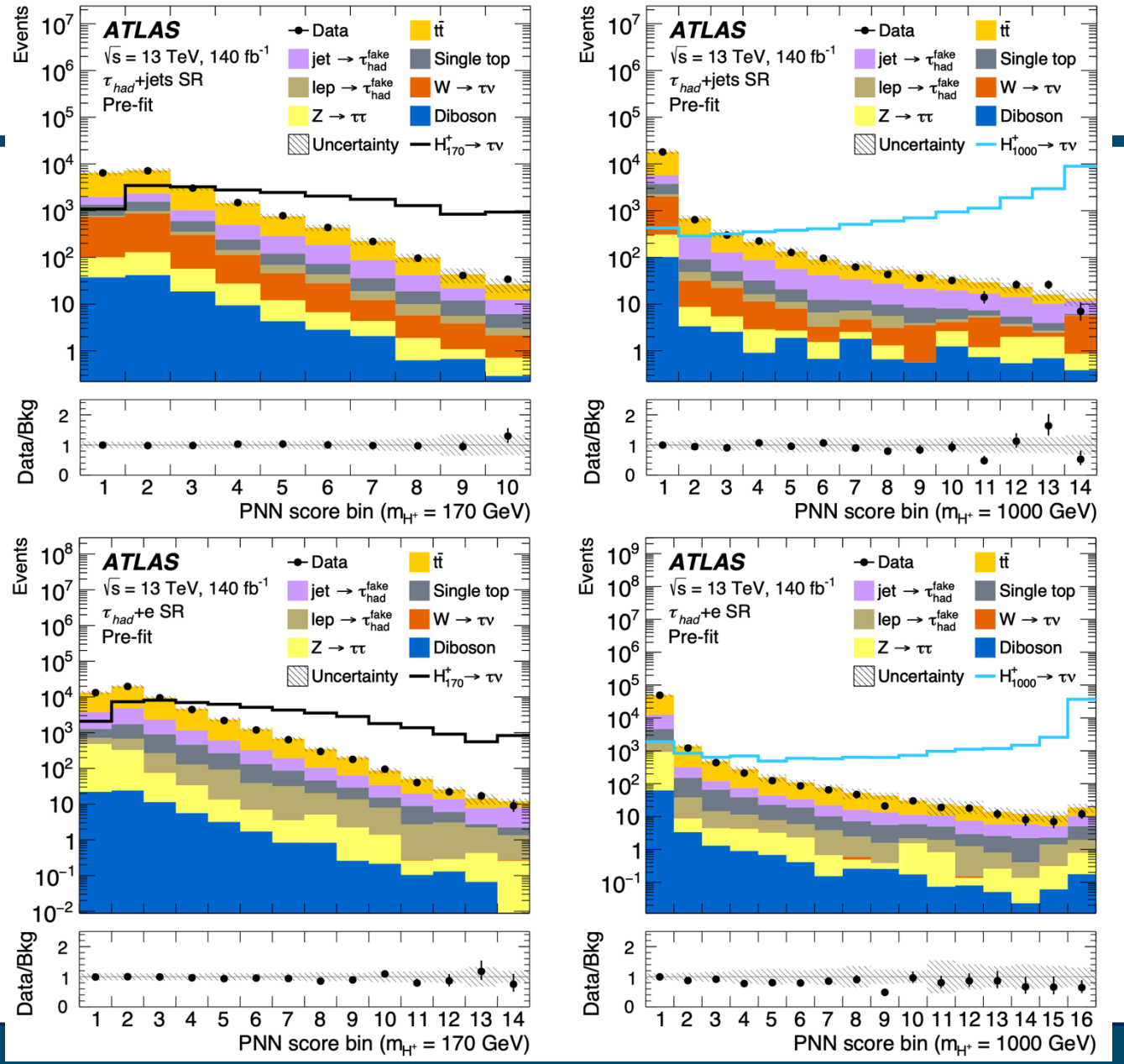
Validation:

- Inputs and output validated in **control regions ($t\bar{t}$, W/Z+jets)**

Pre-fit agreement:

- Data consistent with **SM expectations** before fitting

- Stable performance across **low and high mass hypotheses**



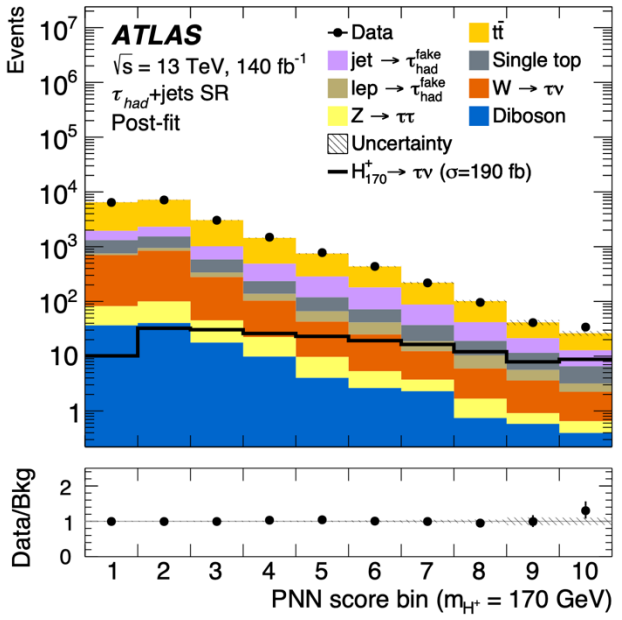
[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector*, Physical Review D 111, 072006 (2025)

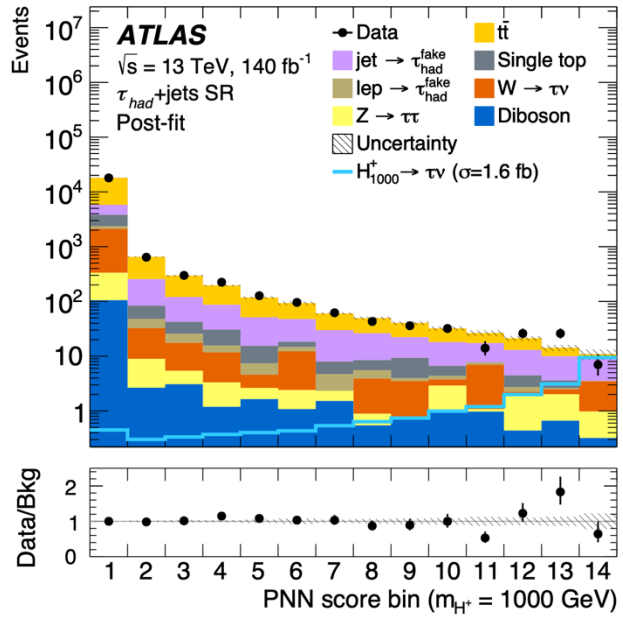
Post-fit results

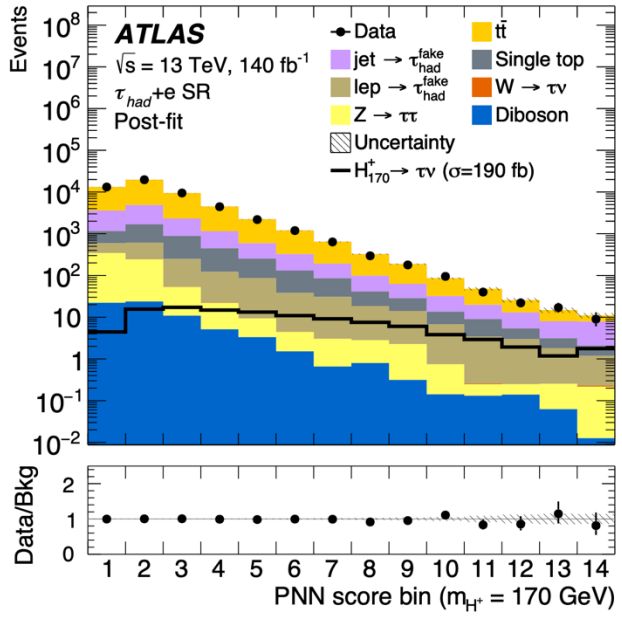
- **Simultaneous likelihood fit** across all signal regions

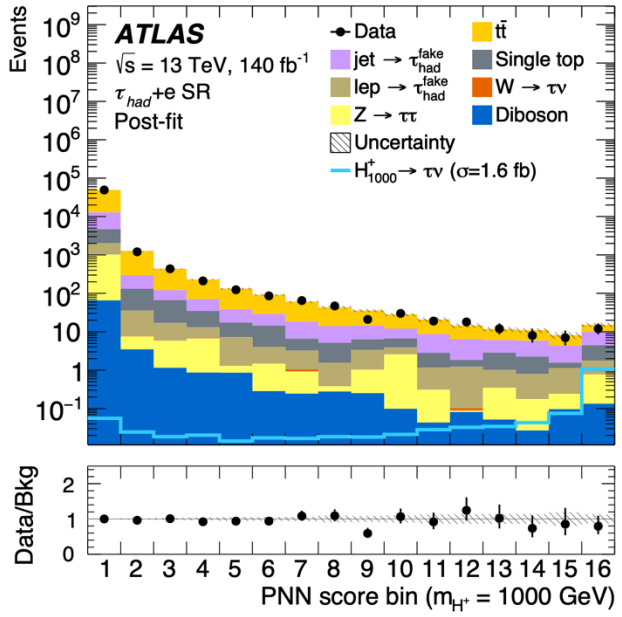
Post-fit agreement:

- Excellent agreement between **data and background model**
- No significant excess observed









[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector*, Physical Review D 111, 072006 (2025)

Systematic uncertainties

Source of systematic uncertainty	Impact on $\sigma(\hat{\mu})$ [%]	
	$m_{H^+} = 170$ GeV	$m_{H^+} = 1000$ GeV
Experimental		
Luminosity	1.0	1.0
Trigger	12	4.1
$\tau_{\text{had-vis}}$	12	8.3
Jet	27	17
Electron	1.9	4.0
Muon	9.3	2.2
$E_{\text{T}}^{\text{miss}}$	35	3.3
Fake-factor method	21	16
Signal and background models		
$t\bar{t}$ modeling	20	14
Single-top-quark modeling	46	2.6
W/Z+jets modeling	13	51
Cross-sections ($W/Z/VV/t$)	2.3	15
H^+ signal modeling	8.5	1.0
Total	84	70

[1] ATLAS Collaboration, *Search for charged Higgs bosons produced in top-quark decays or in association with top quarks and decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in 13 TeV pp collisions with the ATLAS detector,* Physical Review D 111, 072006 (2025)