# Background estimation in the search for $H^{+} \to \tau \nu$ in the ATLAS experiment

Cracow, 28.04.2023

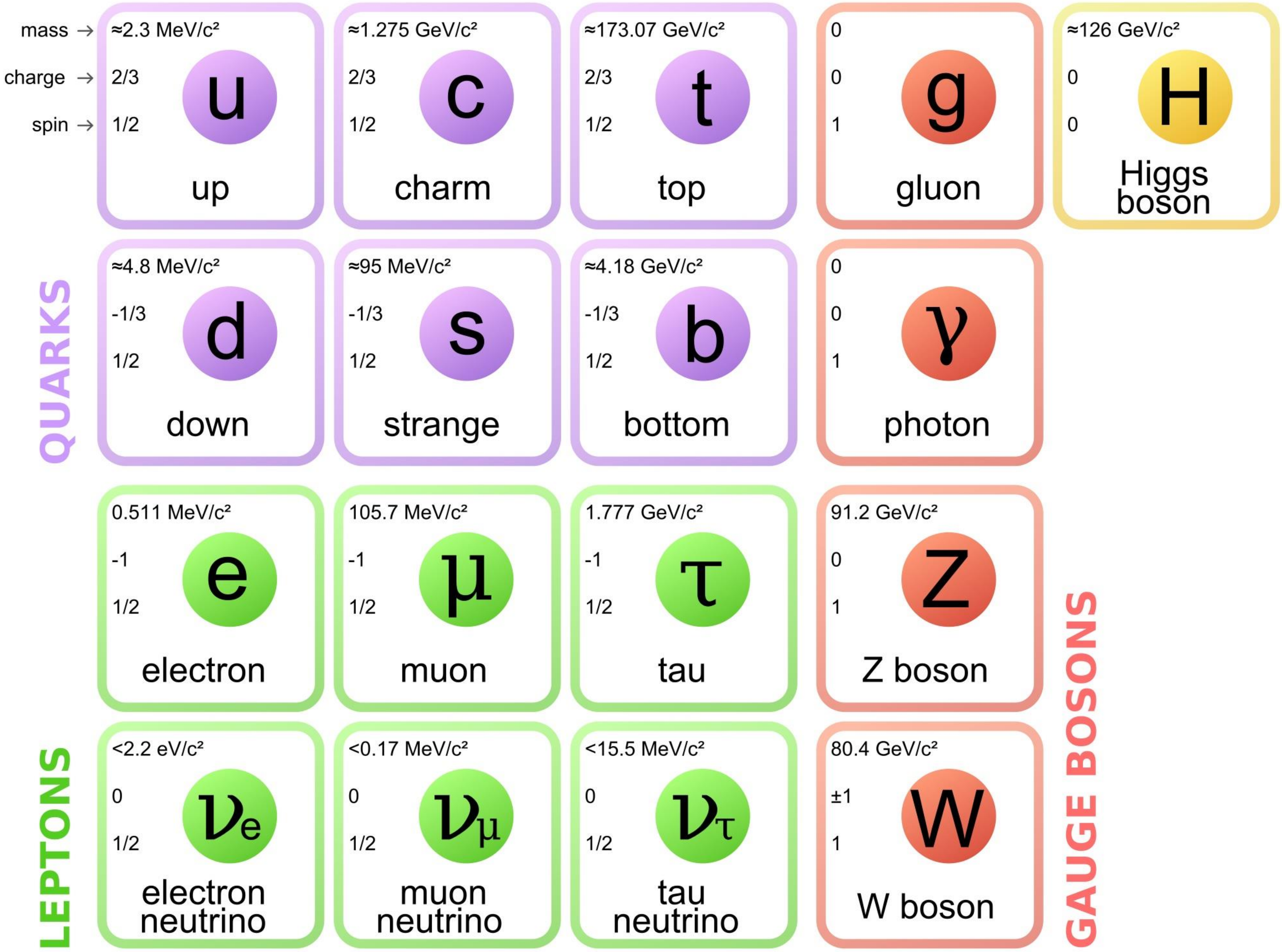

Speaker: Monika Juzek

Supervisor: dr hab. Pawel Bruckman de Renstrom

THE HENRYK NIEWODNICZAŃSKI
INSTITUTE OF NUCLEAR PHYSICS
POLISH ACADEMY OF SCIENCES

ATLAS EXPERIMENT

# World of subatomic particles – the Standard Model



# Beyond the SM – The search for $H^+ \rightarrow \tau \upsilon$



Valid problems of the SM:

- does not inlude gravity,
- dark matter and dark energy,
- neutrino masses etc.

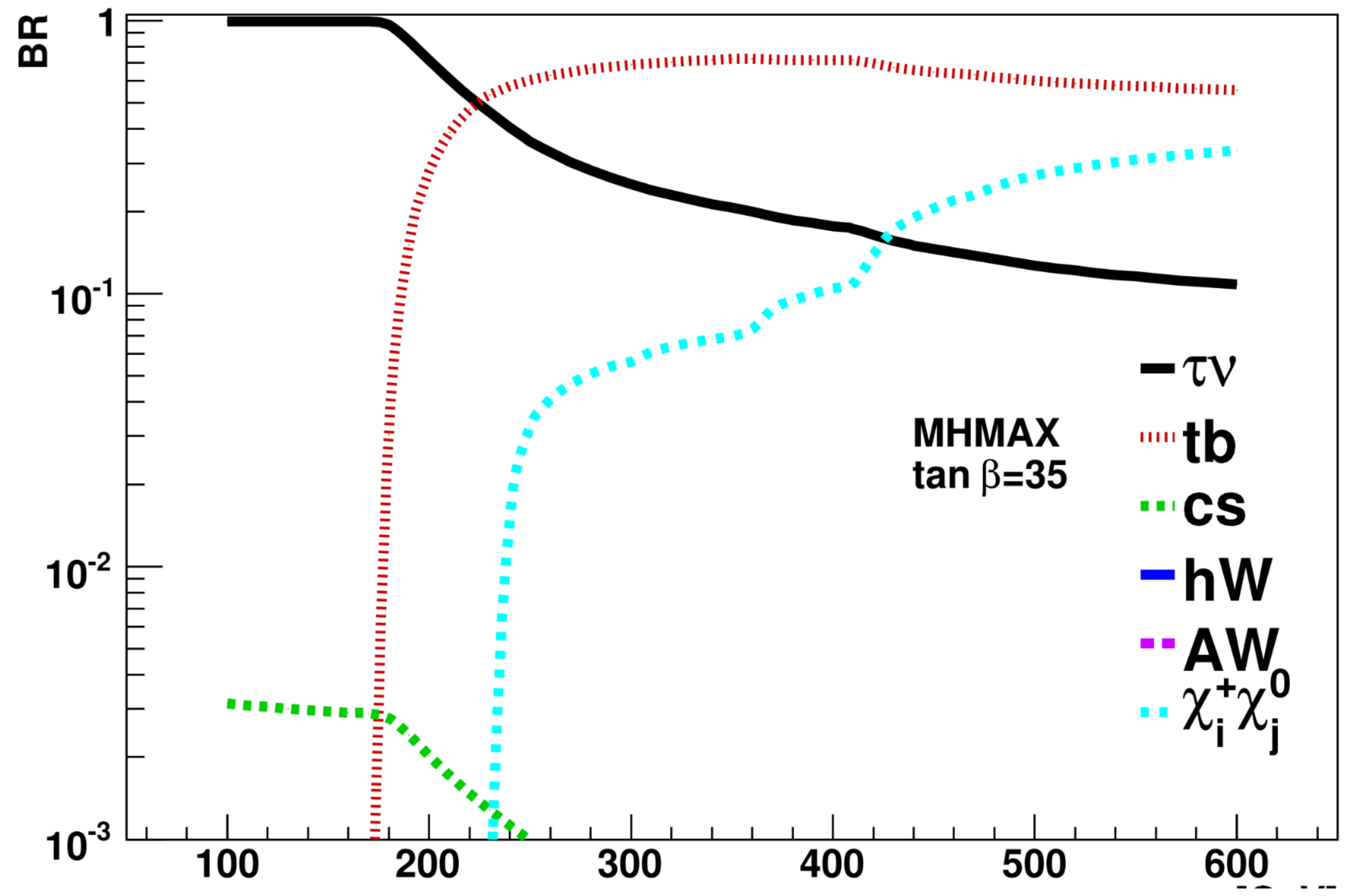
# Beyond the SM – The search for $H^+ \to \tau\upsilon$

**New fundamental theory needed**

↓

**Minimal Supersymmetric Standard Model (MSSM) predicting additional particles like charged Higgs boson**

BR(H+ → ...) - probablity of decay



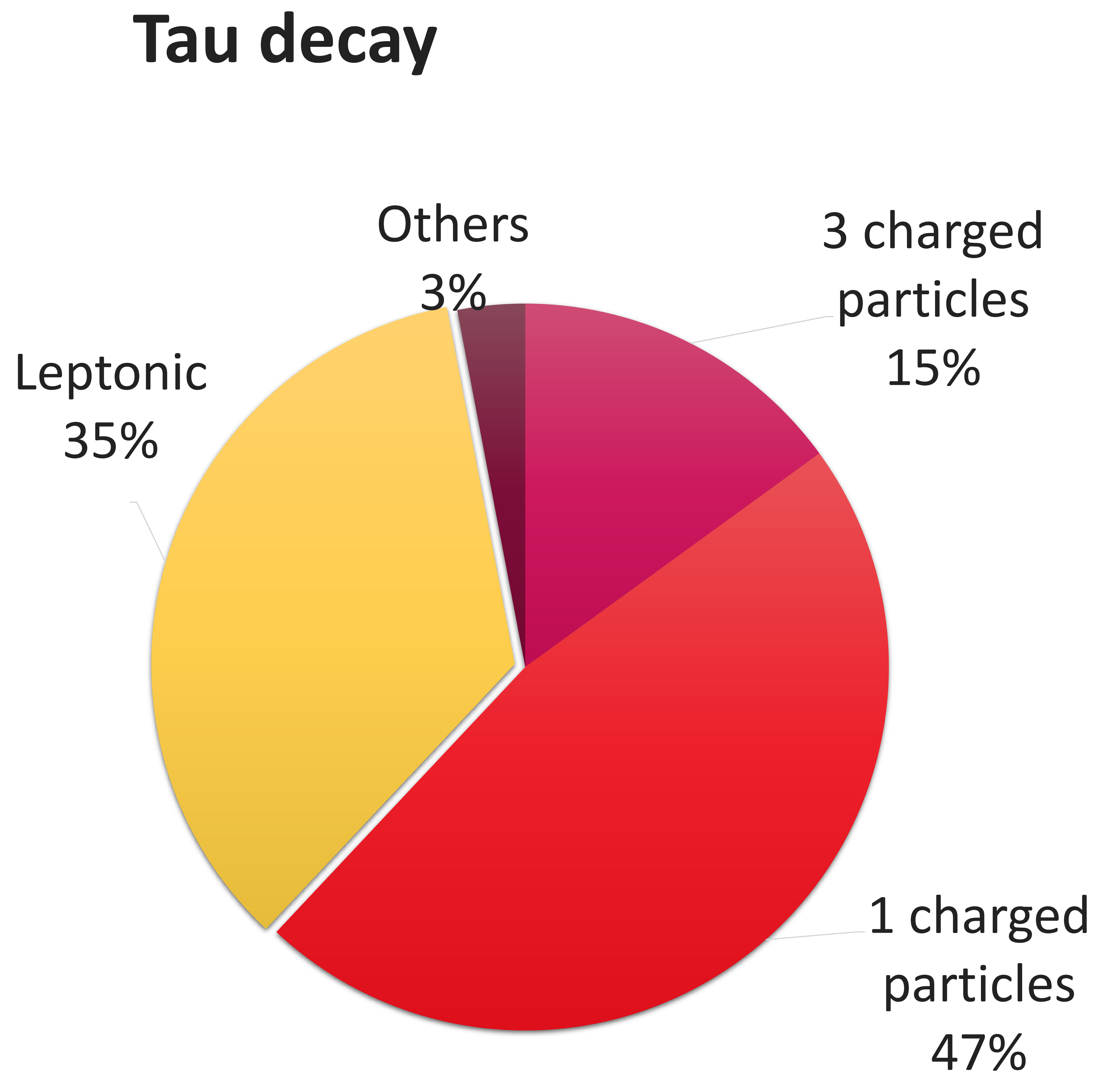

Mass of charged Higgs [GeV]

- **To prove (or disprove) a MSSM theory we're looking for $H^+ \to \tau\upsilon$ process in the ATLAS data collected in 2015-2018**

Previous analysis using 2015-2016 dataset: JHEP 09(2018)139

# Tau lepton in ATLAS analyses

- Tau lepton is a short-lived particle, hence we can observe only its decay products
- It is the heaviest from leptons, so it can decay into mostly **hadrons** (particles made of quarks) or **leptons**
- **Hadronically decaying taus are difficult to identify in data**



# Background estimation - introduction

In particle physics analysis:

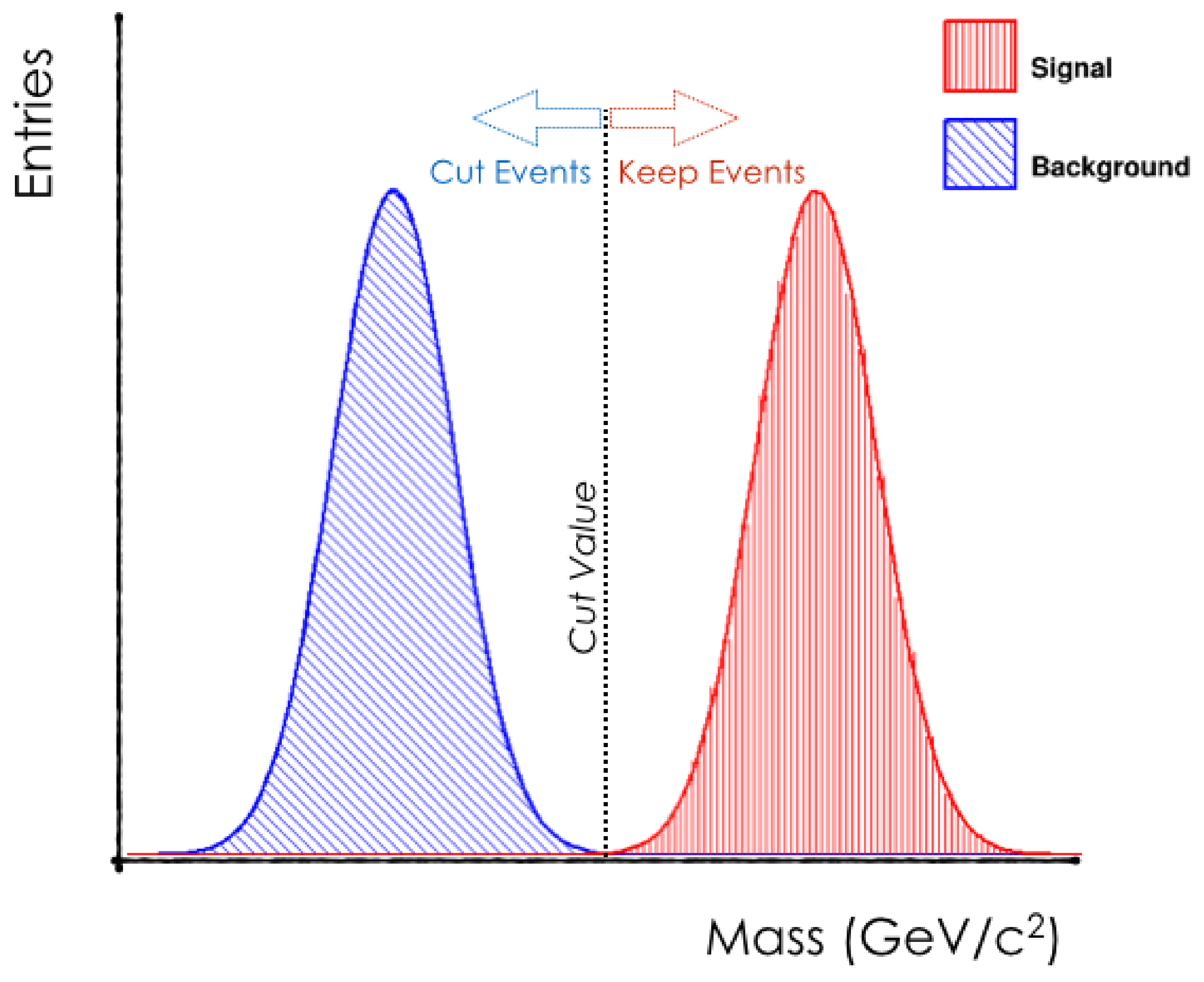
- **”signal”** - whatever particular process we are interested in

$H^{+} \rightarrow \tau\upsilon$

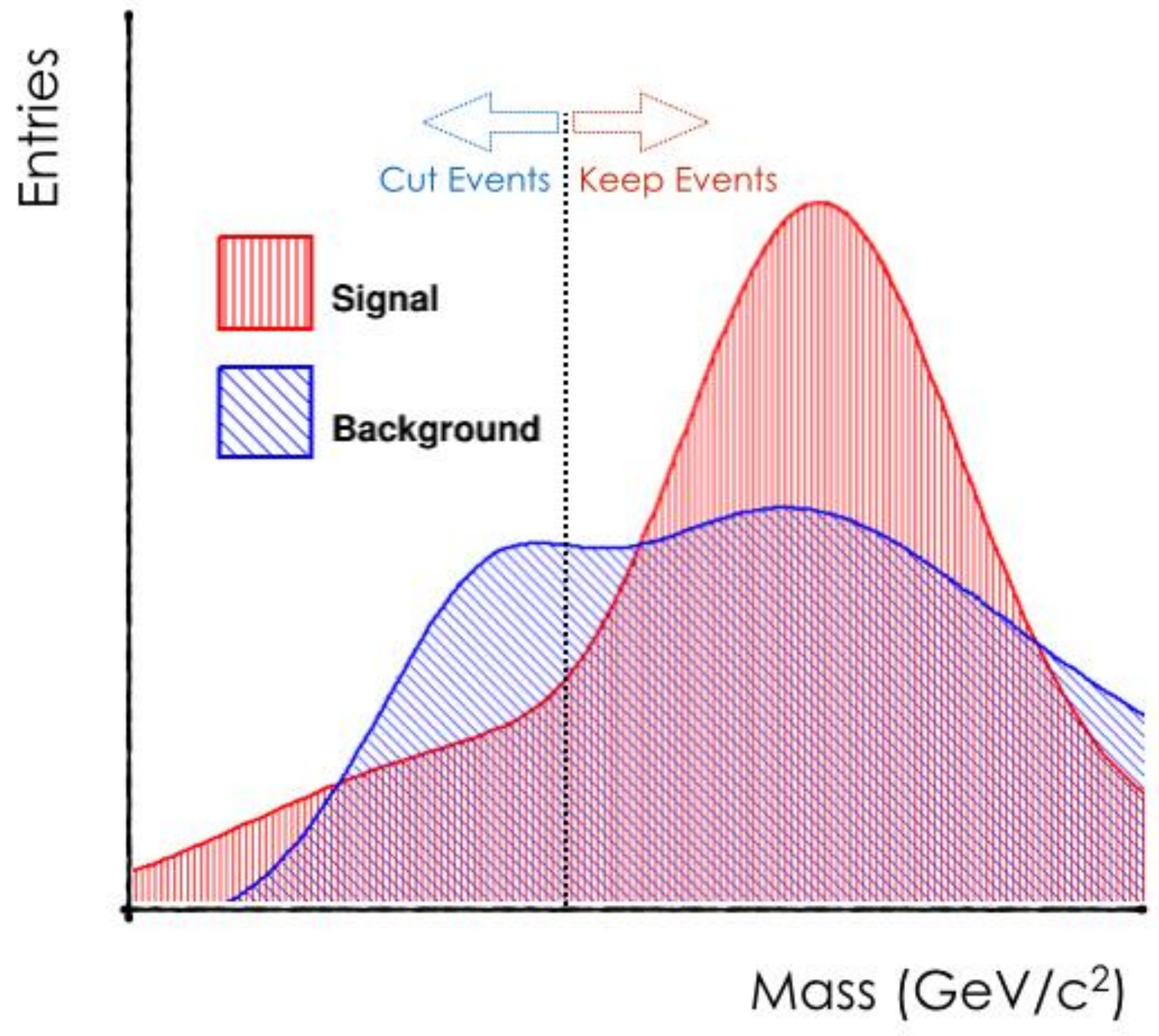
- **„background”** - processes which might look a little bit (or a lot) like the signal process we care about, but are not

**Everything else**

**We want to separate them by using cuts** – set of restrictions on the particle parametrs that would give us the biggest signal part possible (and lowest background)




Perfect world



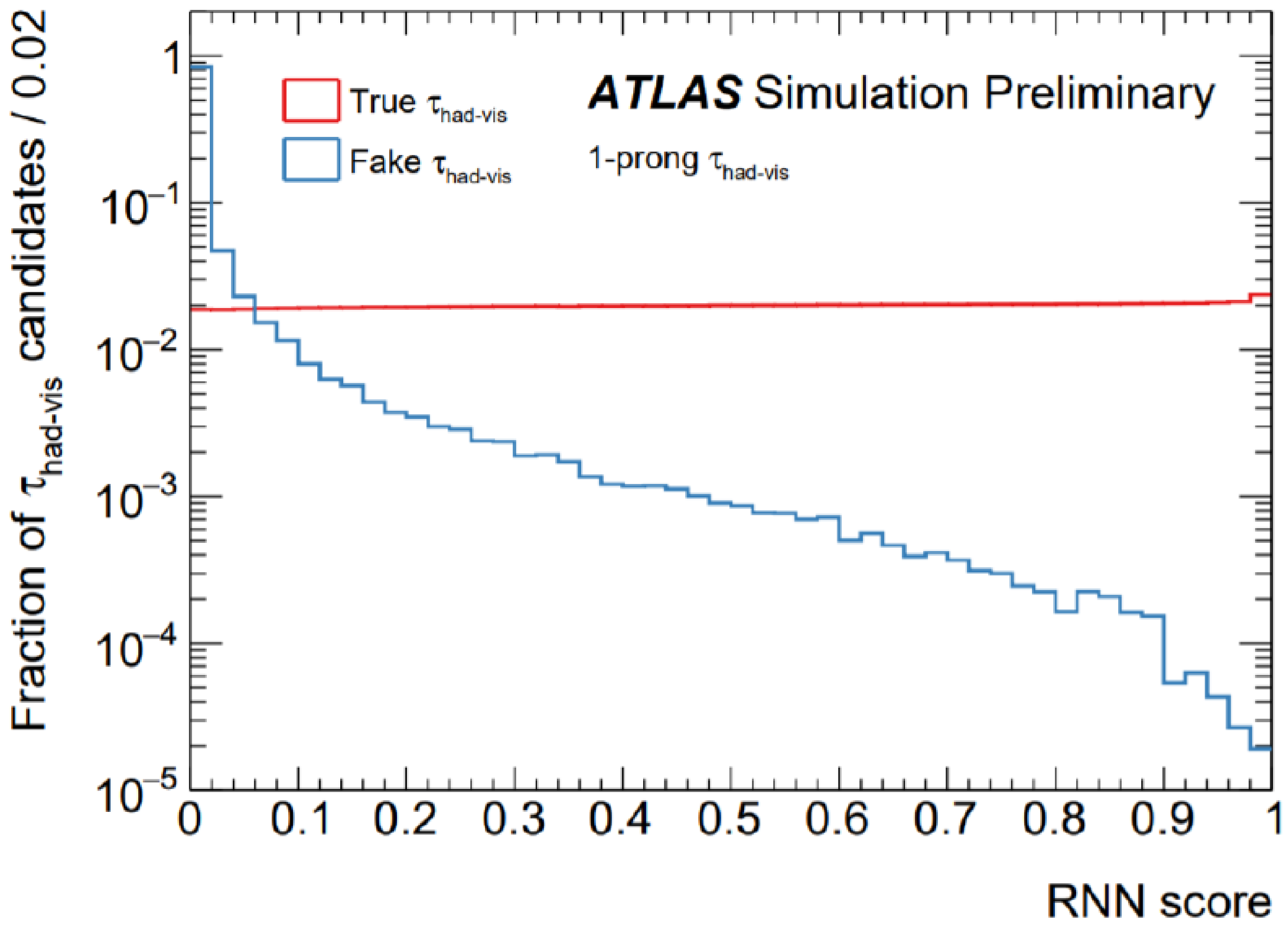

Reality

# Background estimation

To estimate $H^+ \rightarrow \tau \upsilon$ background:

I. **Monte Carlo generators** are used to predict which SM physical processes we can observe during collision ($t\bar{t}$, $Wt$, $V$+jets events)

II. **Machine learning** (Recurrent Neural Networks) for proper tau identification



# Background estimation

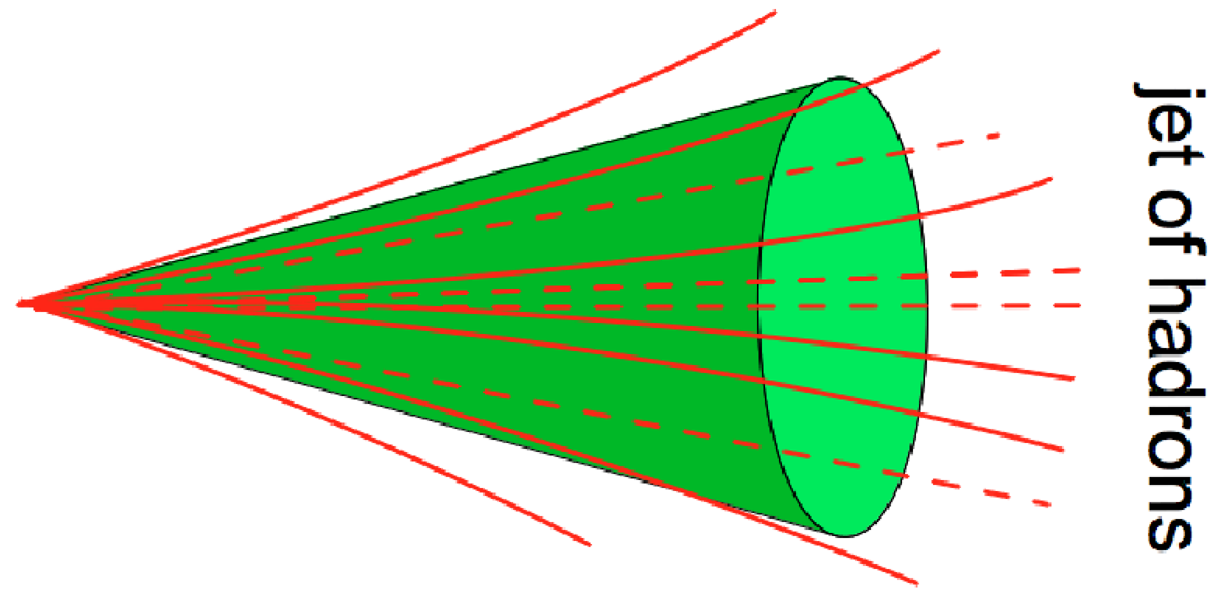
To estimate $H^+ \to \tau\upsilon$ background:

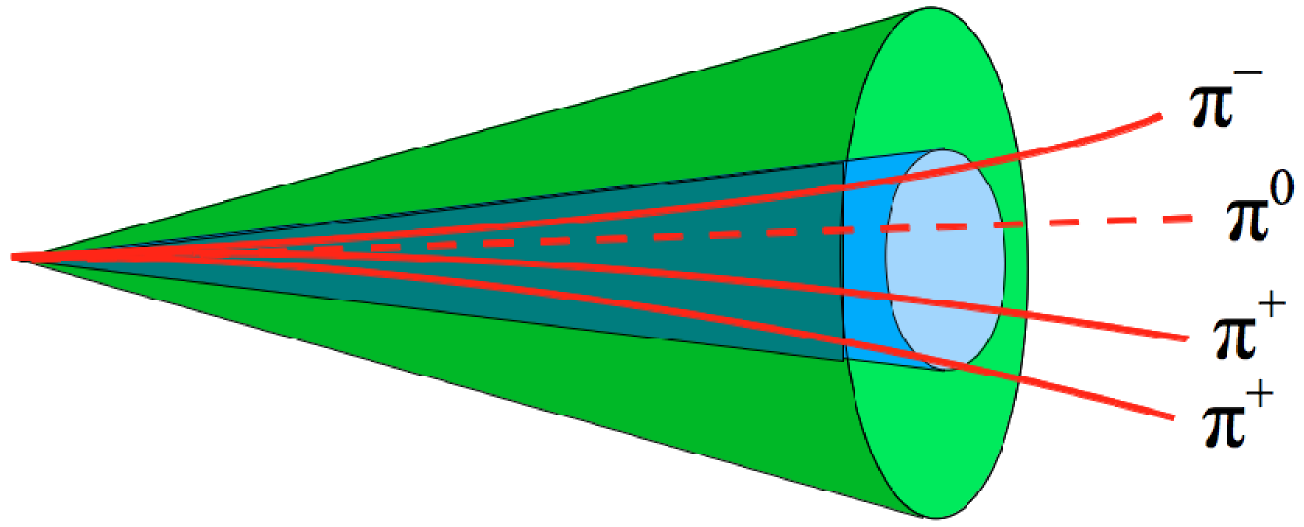
I. M ... hich
SN
co

II. M ... for
proper tau identification

**However, for analysis with tau leptons in final states it is not enough**

**Jet - narrow cone of hadrons and other particles produced during collisions of gluons and quarks**




Tau decay

- **Main source of background comes from the jets misidentified as tau, these are poorly modelled by the Monte Carlo method**

# Fake Factor Method

- Data-driven method of the fake-tau background determination
- The fake-$\tau$ background in SR with the nominal object selection (**$\tau$ ID**) is estimated from events with a reversed selection (**anti-$\tau$ ID**), using **fake factors (FF)** **for certain control region (CR) defined as:**

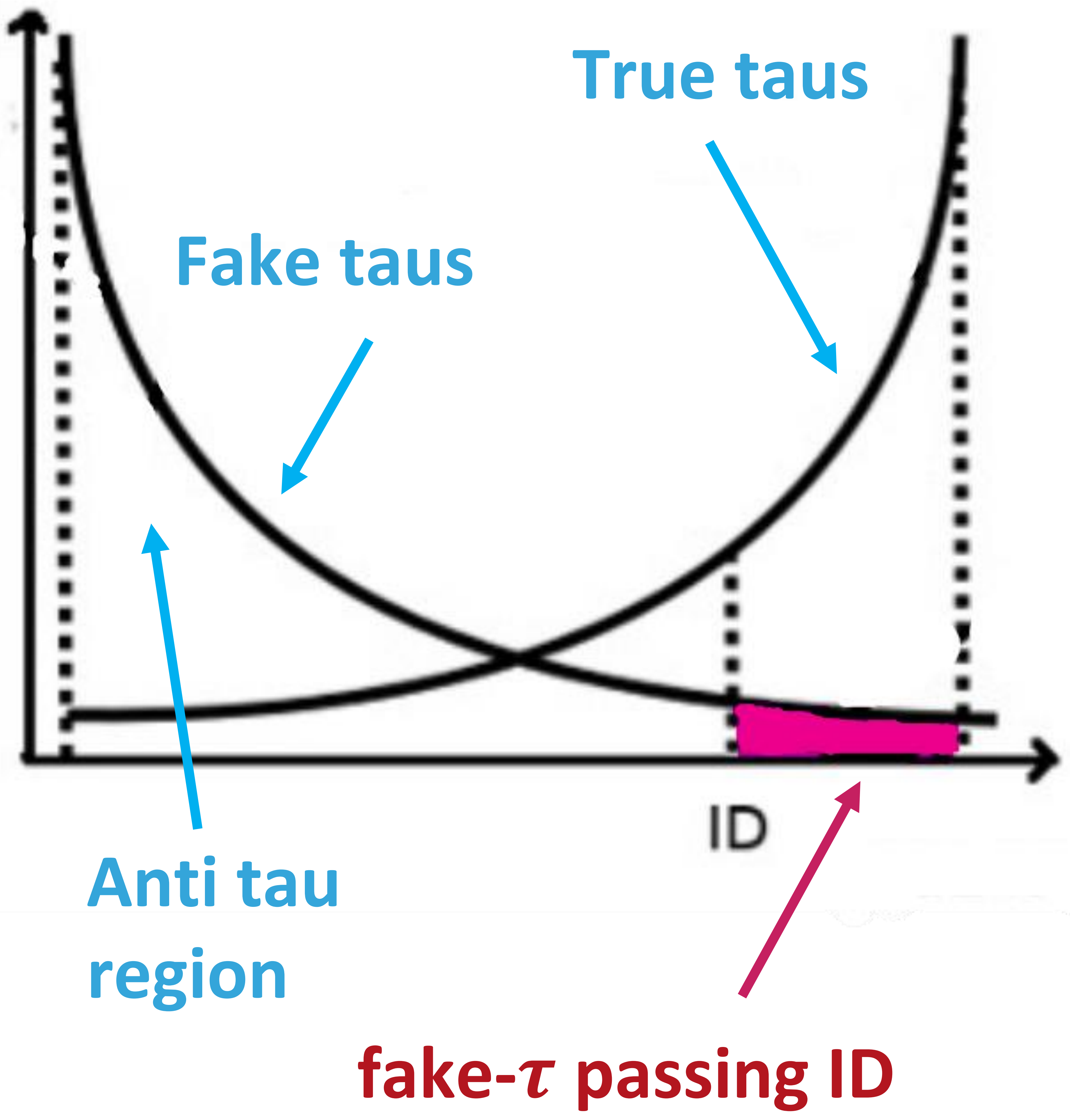
$$FF = \frac{N^{CR}_{\tau-id}}{N^{CR}_{anti\tau-id}}$$

$\longrightarrow$ Events that pass $\tau$ identification

$\longrightarrow$ Events that fail the nominal $\tau$ identification

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

$$N^{\tau}_{\text{fakes}} = N^{\text{SR}}_{\text{anti-}\tau} \times \text{FF}$$



# Fake Factor Method in $H^{+} \rightarrow \tau \upsilon$ analysis

- **Complication: some true taus sneak into the CR and the "SR but fail ID" regions**

- **Solution - subtracting them with the use of MC:**

$$\mathrm{FF} = \frac{N_{\tau}^{\mathrm{CR}}(\mathrm{data}) - N_{\tau}^{\mathrm{CR}}(\mathrm{MC}, \mathrm{true}-\tau)}{N_{\mathrm{anti}-\tau}^{\mathrm{CR}}(\mathrm{data}) - N_{\mathrm{anti}-\tau}^{\mathrm{CR}}(\mathrm{MC}, \mathrm{true}-\tau)}$$

- Parametrization of the FFs : $p_T^{\tau}$ and numer of charged tracks (1-track or 3-track $\tau$)
- **In the $H^{\pm} \rightarrow \tau\nu$ analysis FFs are extracted in two CRs:**
  - I. **Multi-jets CR** → gluon-initiated jets
  - II. **W+jets CR** → quark-initiated jet

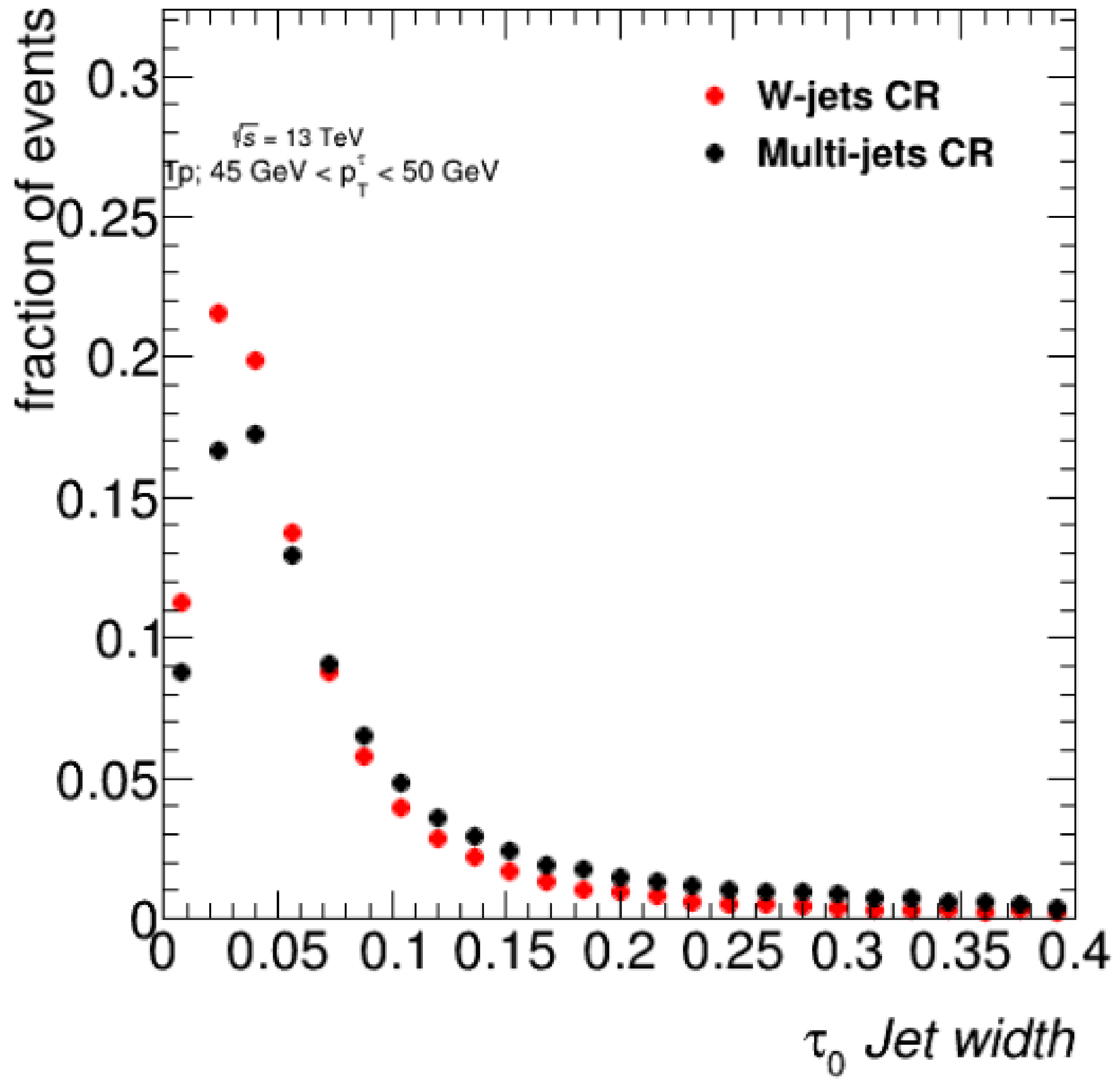
| FF determination Data | FF application Data |
|---|---|
| CR Pass ID | SR Pass ID |
| CR Fail ID | SR but Fail ID |

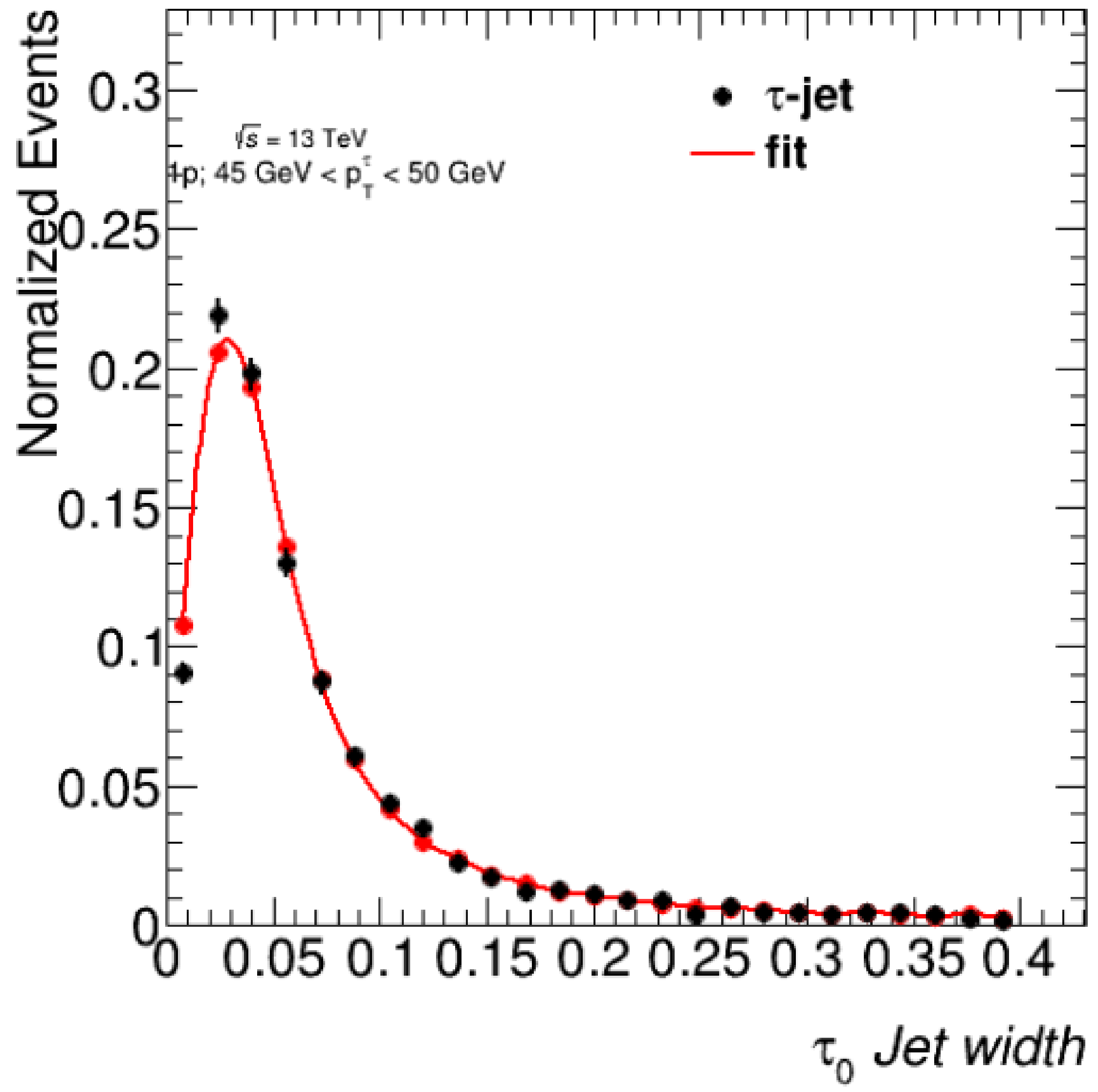
# Quark/gluon composition determination

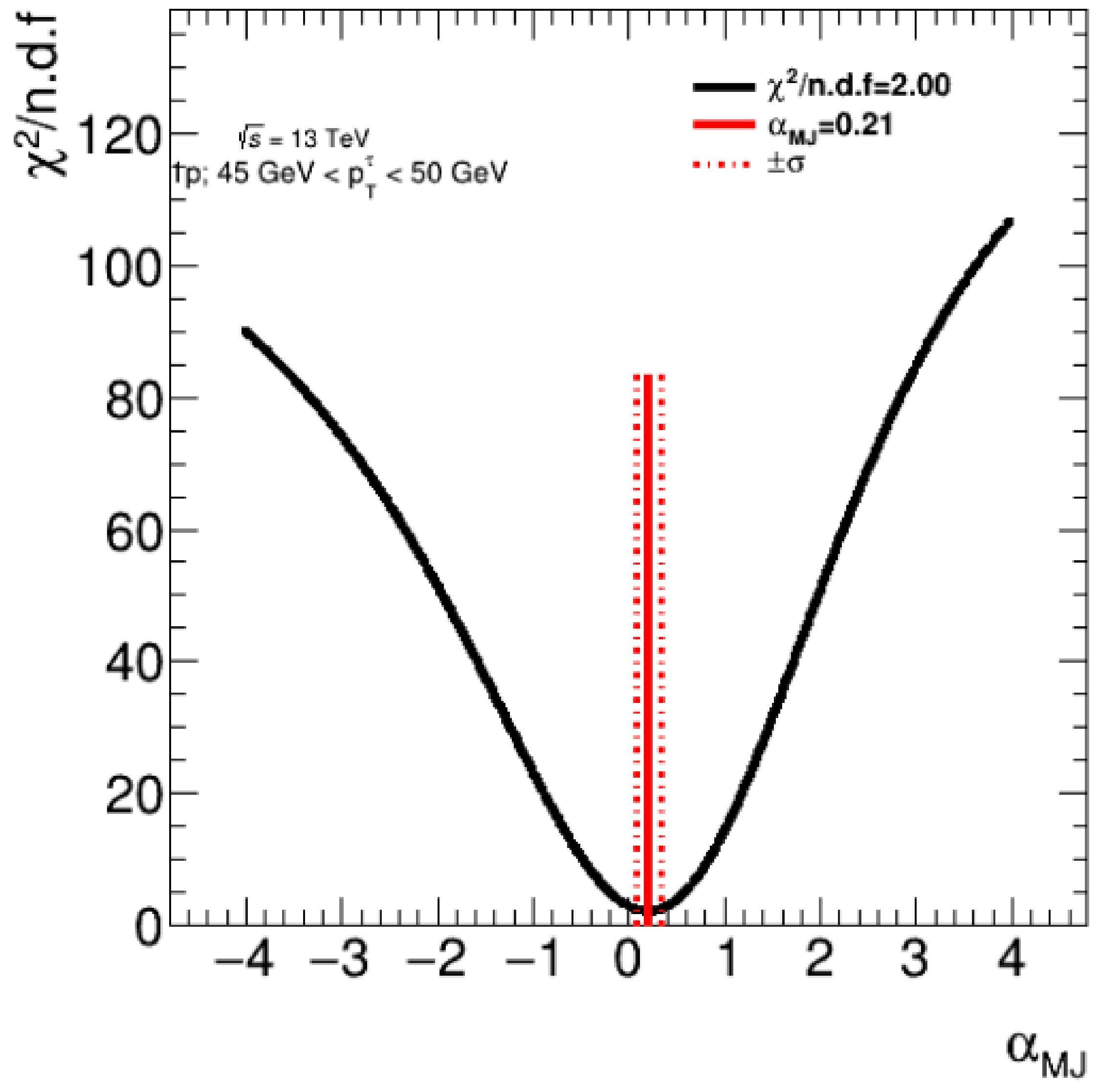
- **To obtain proper jets composition in a signal region, FFs from two CRs are combined using the formula:**

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

- $\boldsymbol{\alpha_{MJ}}$ **parametr is calculated using a template-fit method**:



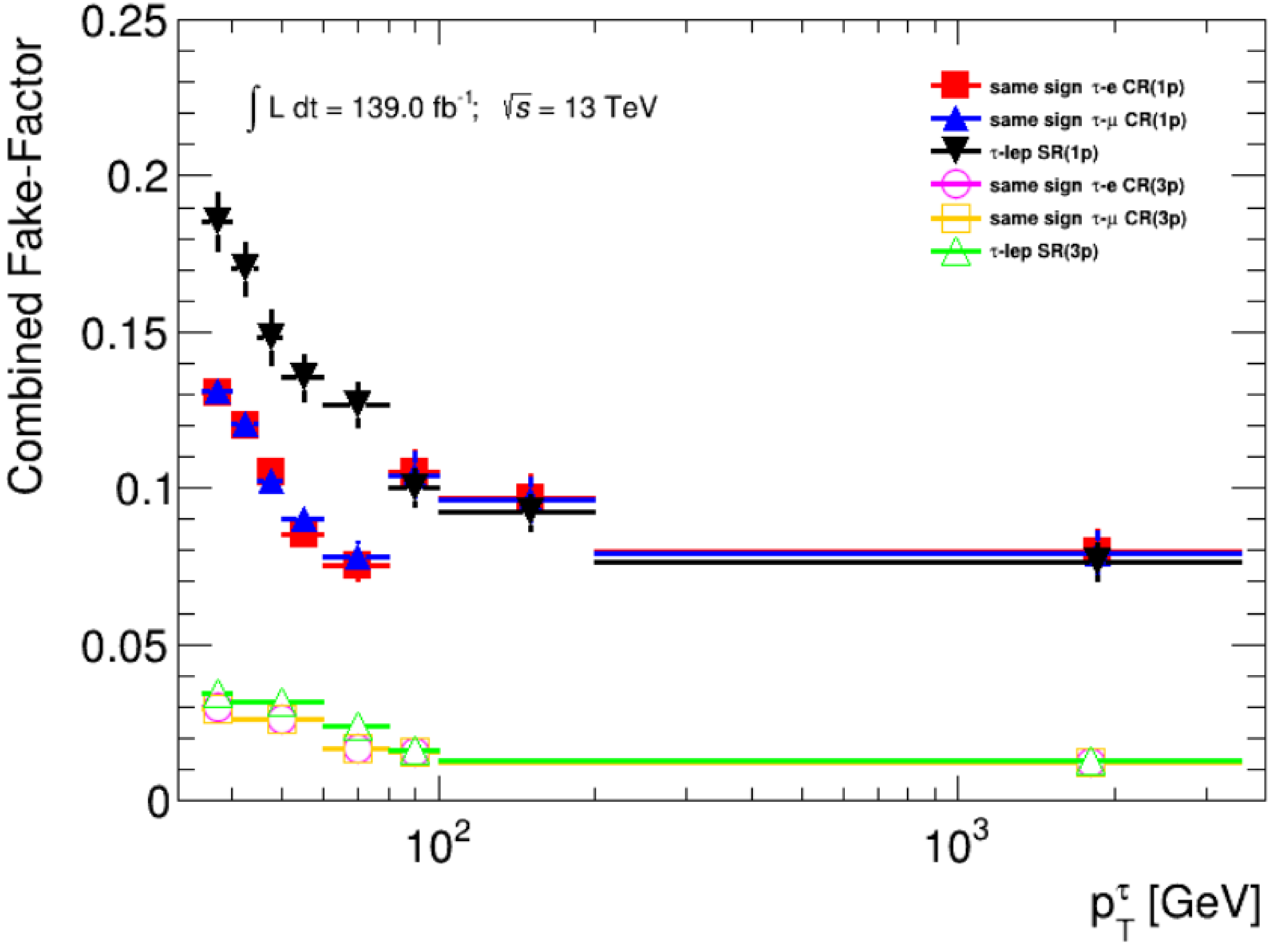



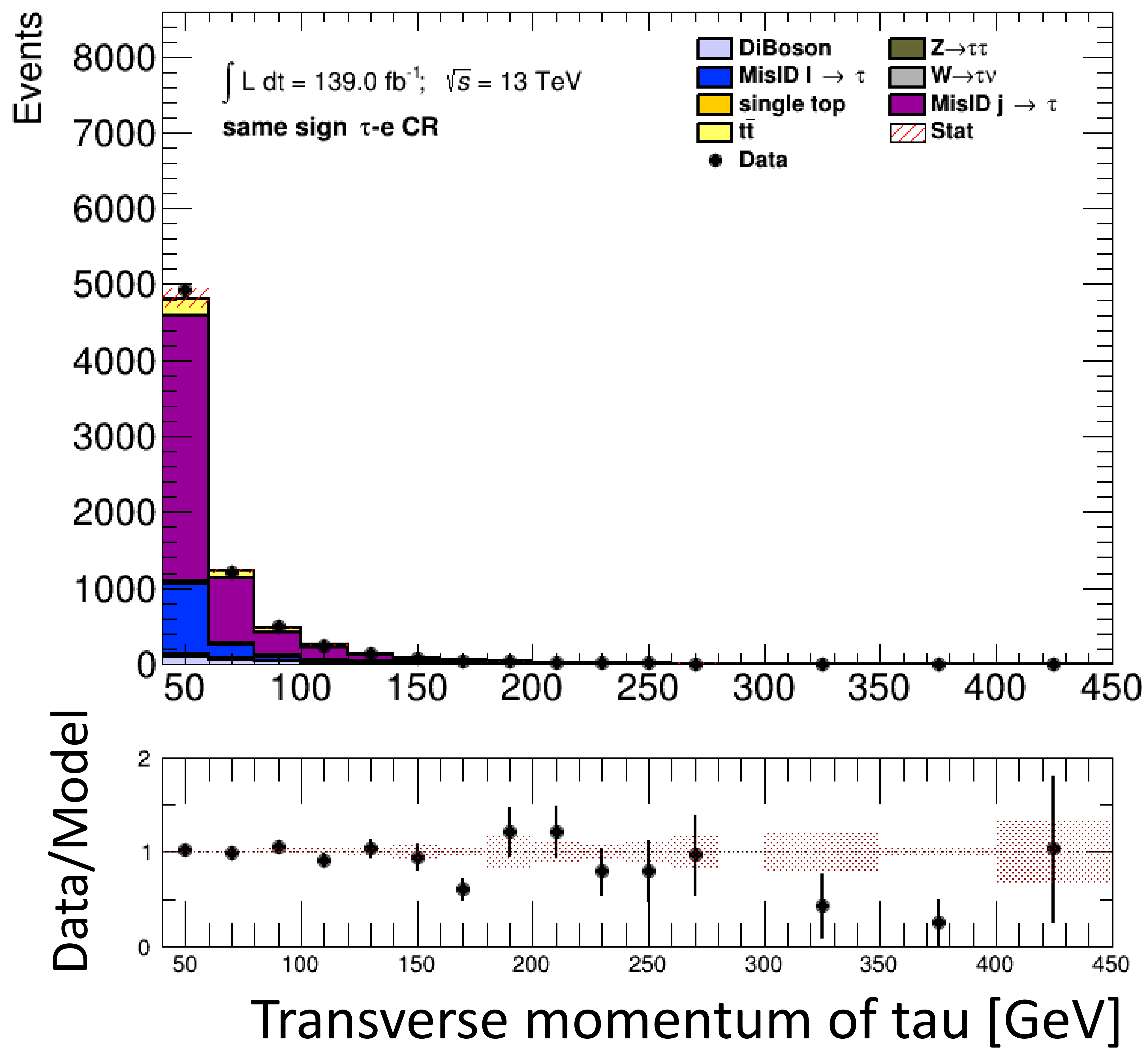

**Creates 2 templates for CRs (distributions for discriminant variable)** → **Then fits it to the SR template**

**$\chi^2$ statistics distribution**

# Validation of the FF method

**τ+lepton channel**



# 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 $\tau$ candidates fulfilling the anti- $\tau$ 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 $\alpha_{MJ}$

V. for the $\Upsilon$ distribution (used in RNN-based $\tau$ ID) the uncertainty of the inverse transform sampling method (Smirnov transformation for $\Upsilon$)

VI. the modelling of heavy-flavour jets mimicking $\tau$ candidates

# Systematics of Fake Factor method in $H^{\pm} \rightarrow \tau\nu$ analysis (2015-2016 data)

| | Source of uncertainty | $\tau_{\text{had-vis}}$+jets Effect on yield | $\tau_{\text{had-vis}}$+lepton Effect on yield |
|---|---|---|---|
| I. | Fake factors: jet composition | 1.6% | 0.6% |
| II. | Fake factors: statistical uncertainties | 1.6% | 1.7% |
| III. | Fake factors: prompt $\tau_{\text{had-vis}}$ in the anti-$\tau$ CR | +5.6% -8.3% | +4.8% -7.2% |
| IV. | Fake factors: $\alpha_{\text{MJ}}$ uncertainty | 7% | 6.2% |
| V. | Fake factors: Smirnov transform. | 0% | 0% |
| VI. | Fake factors: heavy flavor jet fraction. | 5% | 5% |

Paper: JHEP 09(2018)139

- Estimation of the fake jets $\rightarrow \tau$ background is the main source of systematic uncertainties in the low- and intermediate-mass $H^{\pm}$ search and the second major source (after the signal modelling) for large $H^{\pm}$ masses

# Summary

- **One of the key roles of the ATLAS experiment is to search for physics beyond the Standard Model**

- **Data-driven Fake Factor method for fake $\tau$ leptons background estimation is an important part of the $H^{\pm} \rightarrow \tau\nu$ analysis**. Optimization of the method and detailed uncertainties estimation are essential as a major source of systematic uncertainties

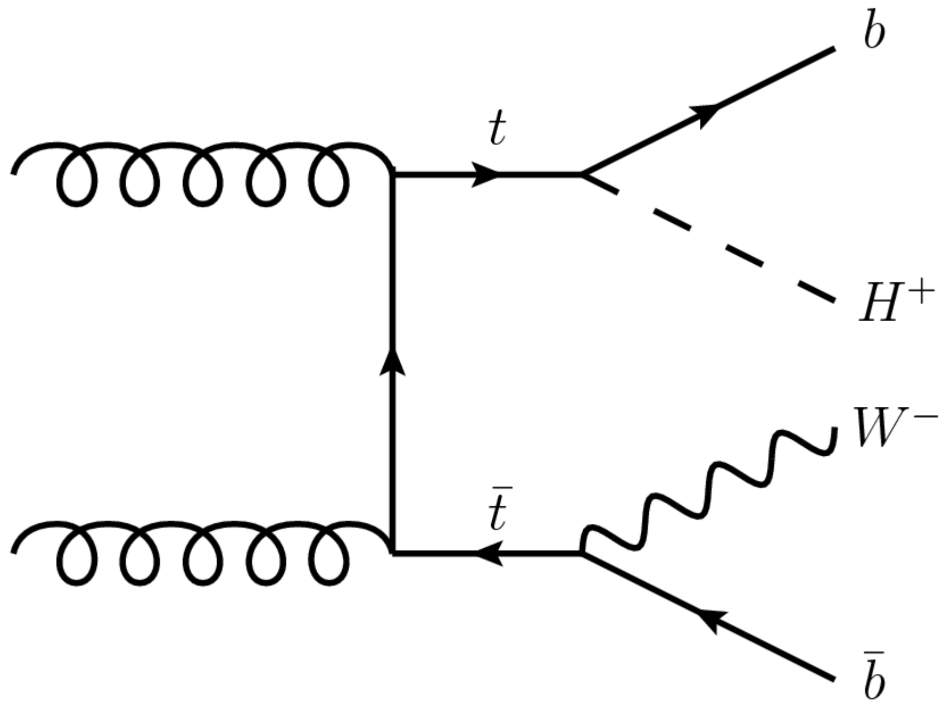
- Currently getting ready to publish results with full Run 2 (2015-2018) data with improved analysis strategy

Thank you for your attention!

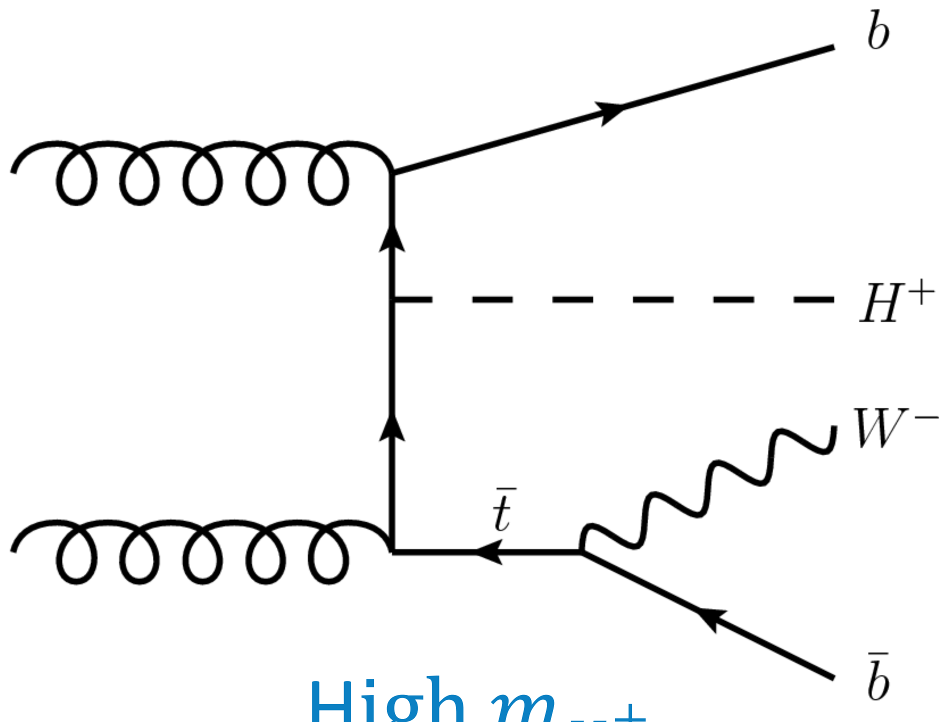
# BACKUP

# $H^{\pm} \rightarrow \tau\nu$ search

- $H^{\pm}$ bosons are predicted in different extensions of the SM that add a second doublet or triplets to the Higgs scalar sector (2HDM, NMSSM, Triplet, etc.)
- Investigated signal mass range: 80 GeV ≤ $m_{H^{\pm}}$ ≤ 3000 GeV
- $H^{\pm}$ production process depends on the mass range but is typically associated with $t$ and $b$ quarks
- 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
- The search using 2015+2016 data was published in: JHEP 09(2018)139




Low & intermediate $m_{H^{\pm}}$




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

# Analysis Overview

- **Search for charged $H^{\pm}$ decaying to τν using 139 fb$^{-1}$ of $pp$ data at $\sqrt{s}$ = 13 TeV collected by the ATLAS experiment at the LHC during full Run-2 period**
- **Recurrent Neural Networks (RNN) based $\tau$ identification** → identification efficiency improved in the range 75–100% compared to one based on Boosted Decision Trees
- Parametrized Neural Networks (PNN) are trained to separate signal from all backgrounds (more in Zuzanna's talk)
- Backgrounds with prompt hadronic $\tau$ :
  - $t\bar{t}$, $V$+jets and diboson events: estimated with MC
- Backgrounds with fake $\tau$:
  - fake lepton → $\tau$ : estimated with MC
  - **fake jets → $\tau$ : estimated with data-driven Fake Factor (FF) method**

ATLAS EXPERIMENT

# Event Selection & MVA Strategy

## τ+jets channel

- 1 medium $\tau$ candidate with $p_{\mathrm{T}}^{\tau}$ > 40 GeV
- no loose leptons ($e$ or $\mu$) with $p_T$ > 20 GeV
- ≥ 3 jets with $p_T$ > 25 GeV, of which at least one is $b$-tagged
- $E_{\mathrm{T}}^{\mathrm{miss}}$ > 150 GeV
- $m_{\mathrm{T}}$ > 50 GeV

## τ+lepton channel

- Exactly 1 $\ell$ ($e$ or $\mu$), with $p_{\mathrm{T}}^{\ell}$ > 30 GeV
- Exactly 1 medium $\tau$ with no additional loose or tighter $\tau$ leptons in the event
- $\ell$ and $\tau$ with opposite signs
- $E_{\mathrm{T}}^{\mathrm{miss}}$ > 50 GeV
- ≥ 1 $b$-tagged jets

$$m_{\mathrm{T}} = \sqrt{2 p_{\mathrm{T}}^{\tau} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos \Delta\phi_{\tau,\mathrm{miss}})},$$

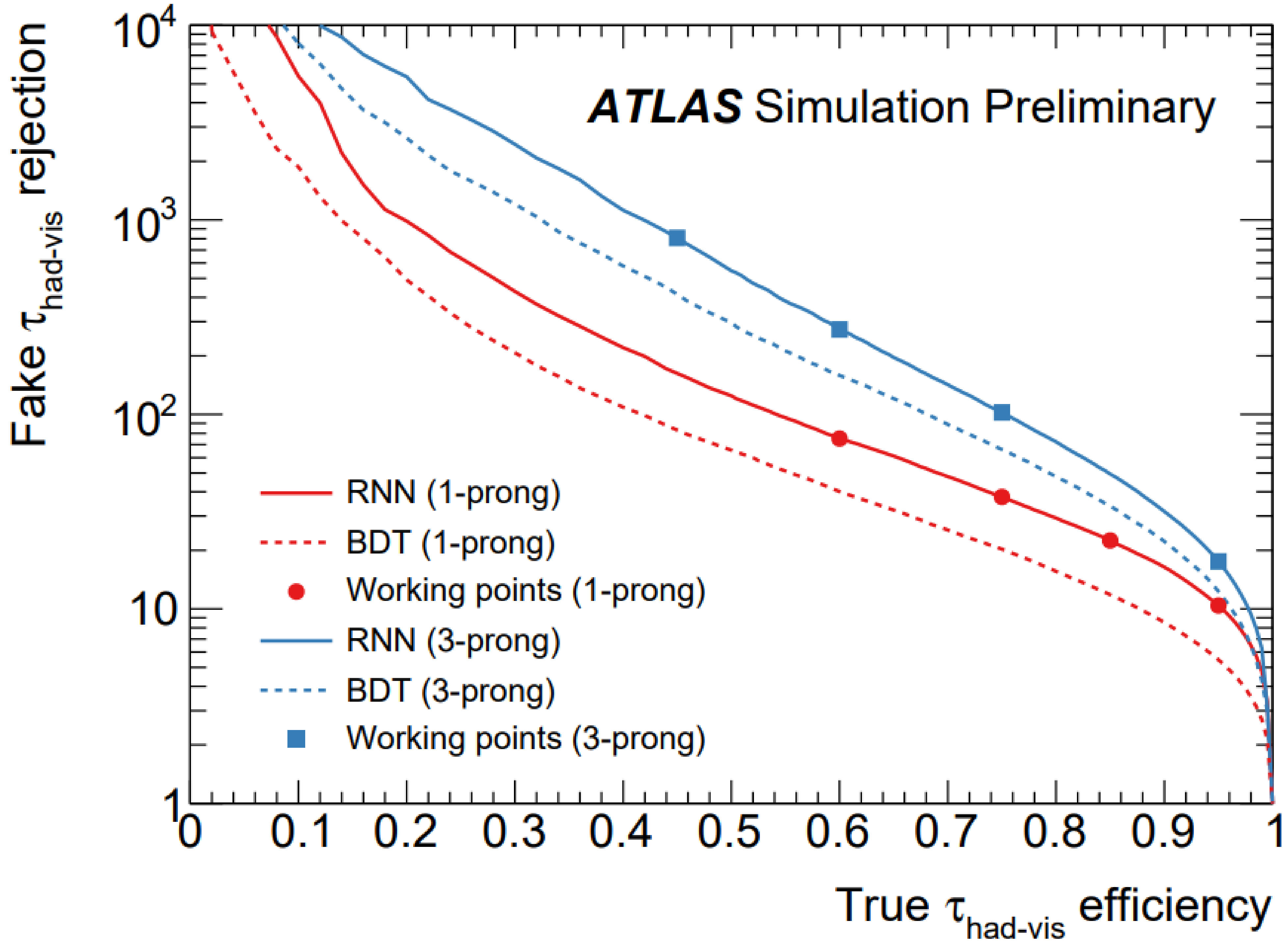
$\Delta\phi_{\tau,miss}$ - the azimuthal angle between the $\tau$ candidate and the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$

- **New multivariate analysis technique, i.e. Parametrized Neural Networks (PNN), previously Boosted Decision Trees (BDT) used**
- PNN score used as a final discriminant in a likelihood fit

ATLAS EXPERIMENT

# Tau identification using RNN

A multivariant discriminanat is a common tool for background estimation, which enables to differentiate correctly reconstructed tau leptons from the other objects. Right now a recurrent neural network (RNN) is used.




| | Observable | 1-prong | 3-prong |
|---|---|---|---|
| Track inputs | $p_T^{\text{seed jet}}$ | ● | ● |
| | $p_T^{\text{track}}$ | ● | ● |
| | $\Delta\eta^{\text{track}}$ | ● | ● |
| | $\Delta\phi^{\text{track}}$ | ● | ● |
| | $\lvert d_0^{\text{track}}\rvert$ | ● | ● |
| | $\lvert z_0^{\text{track}} \sin\theta\rvert$ | ● | ● |
| | $N_{\text{IBL hits}}$ | ● | ● |
| | $N_{\text{Pixel hits}}$ | ● | ● |
| | $N_{\text{SCT hits}}$ | ● | ● |
| Cluster inputs | $p_T^{\text{jet seed}}$ | ● | ● |
| | $E_T^{\text{cluster}}$ | ● | ● |
| | $\Delta\eta^{\text{cluster}}$ | ● | ● |
| | $\Delta\phi^{\text{cluster}}$ | ● | ● |
| | $\lambda_{\text{cluster}}$ | ● | ● |
| | $\langle\lambda^2_{\text{cluster}}\rangle$ | ● | ● |
| | $\langle r^2_{\text{cluster}}\rangle$ | ● | ● |
| High-level inputs | $p_T^{\text{uncalibrated}}$ | ● | ● |
| | $f_{\text{cent}}$ | ● | ● |
| | $f^{-1}_{\text{leadtrack}}$ | ● | ● |
| | $\Delta R_{\text{max}}$ | ● | ● |
| | $\lvert S_{\text{leadtrack}}\rvert$ | ● | |
| | $S_T^{\text{flight}}$ | | ● |
| | $f_{\text{iso}}^{\text{track}}$ | ● | ● |
| | $f_{\text{track}}^{\text{EM}}$ | ● | ● |
| | $p_T^{\text{EM+track}}/p_T$ | ● | ● |
| | $m^{\text{EM+track}}$ | ● | ● |
| | $m^{\text{track}}$ | | ● |

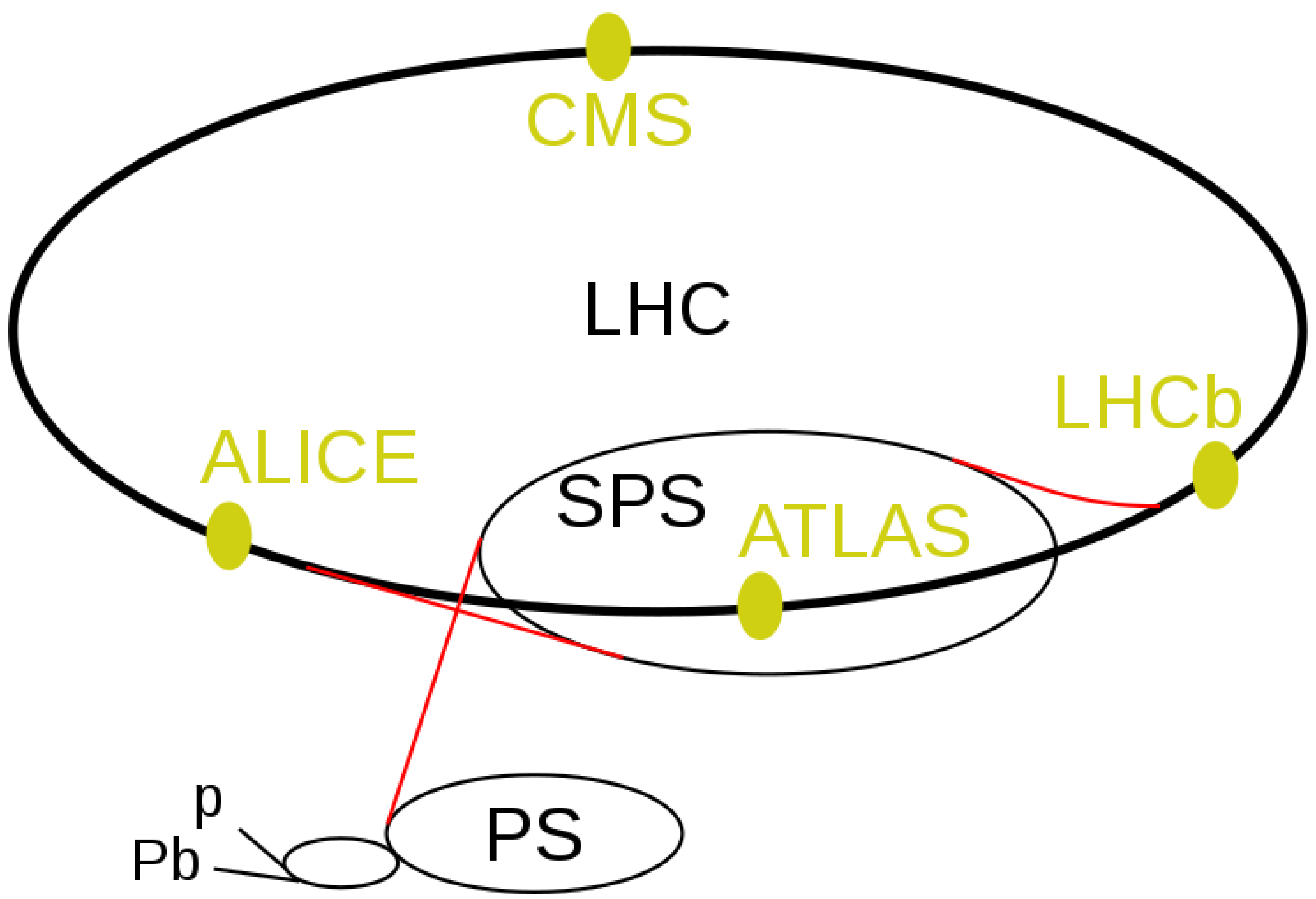
# Definitions of control regions used in FF method

| Multi-jets CR | W+jets CR |
|---|---|
| at least one reconstructed $\tau_{\mathrm{had}}$ candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV | at least one reconstructed $\tau_{\mathrm{had}}$ candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV |
| number of jets 1 | one lepton (electron or muon) |
| $b$-jet veto, electron and muon veto | $b$-jet veto |
| $E_{\mathrm{T}}^{\mathrm{miss}} < 80$ GeV | $p_{\mathrm{T}}$ of electron or muon $> 30$ GeV |
| $m_{\mathrm{T}}(\tau, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}) > 50$ GeV | $60 < m_{\mathrm{T}}(l, E_{\mathrm{T}}^{\mathrm{miss}}) < 160 GeV$ |
| transformed $\tau_{\mathrm{had}}$ BDT score $> 0.02$ | transformed $\tau_{\mathrm{had}}$ BDT score $> 0.02$ |

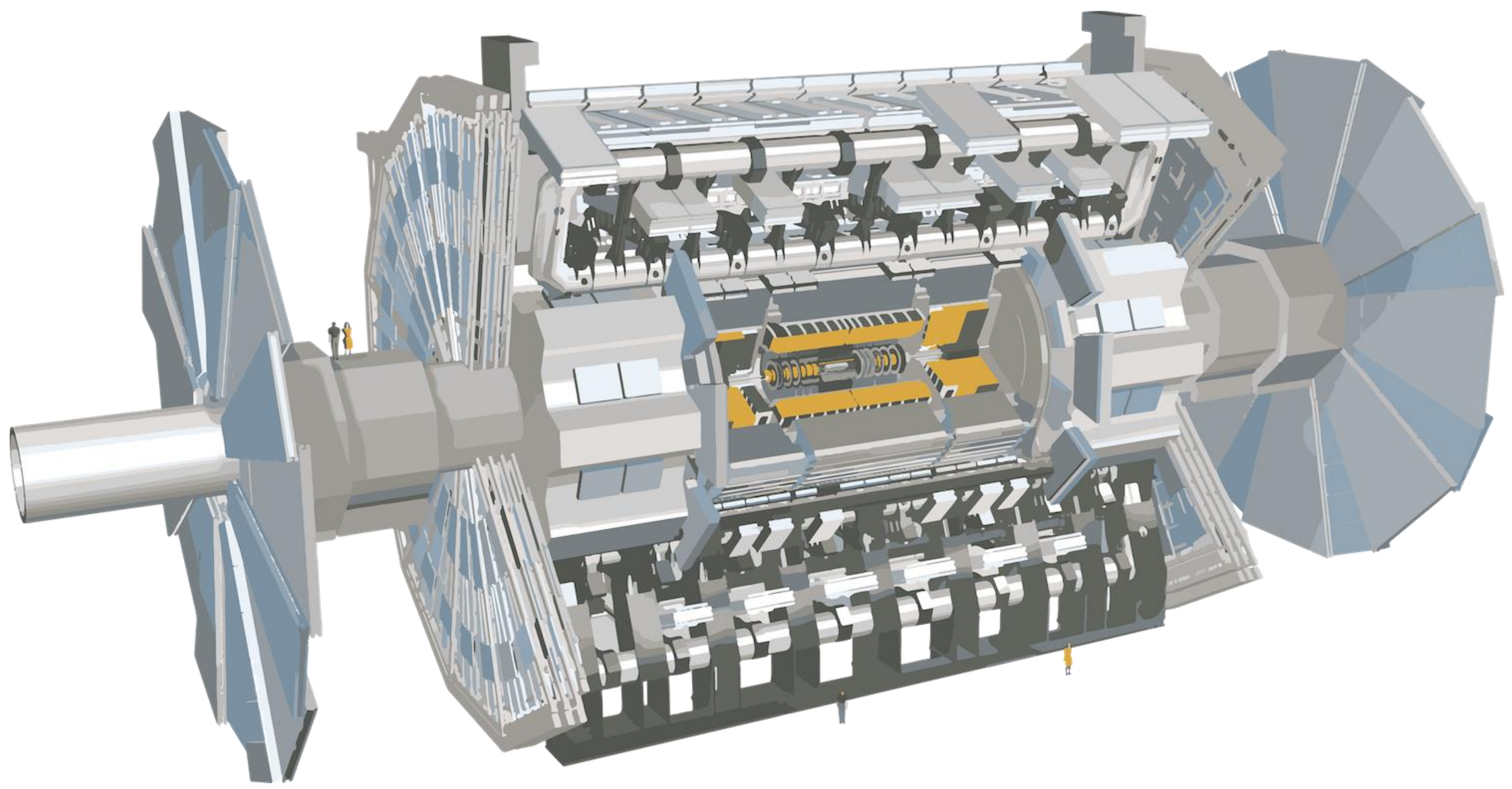
ATLAS EXPERIMENT

# The ATLAS experiment

ATLAS is the largest particle detector experiment at the Large Hadron Collider (LHC), a particle accelerator at CERN in Geneva.




CERN accelerator complex



The ATLAS detector

ATLAS EXPERIMENT