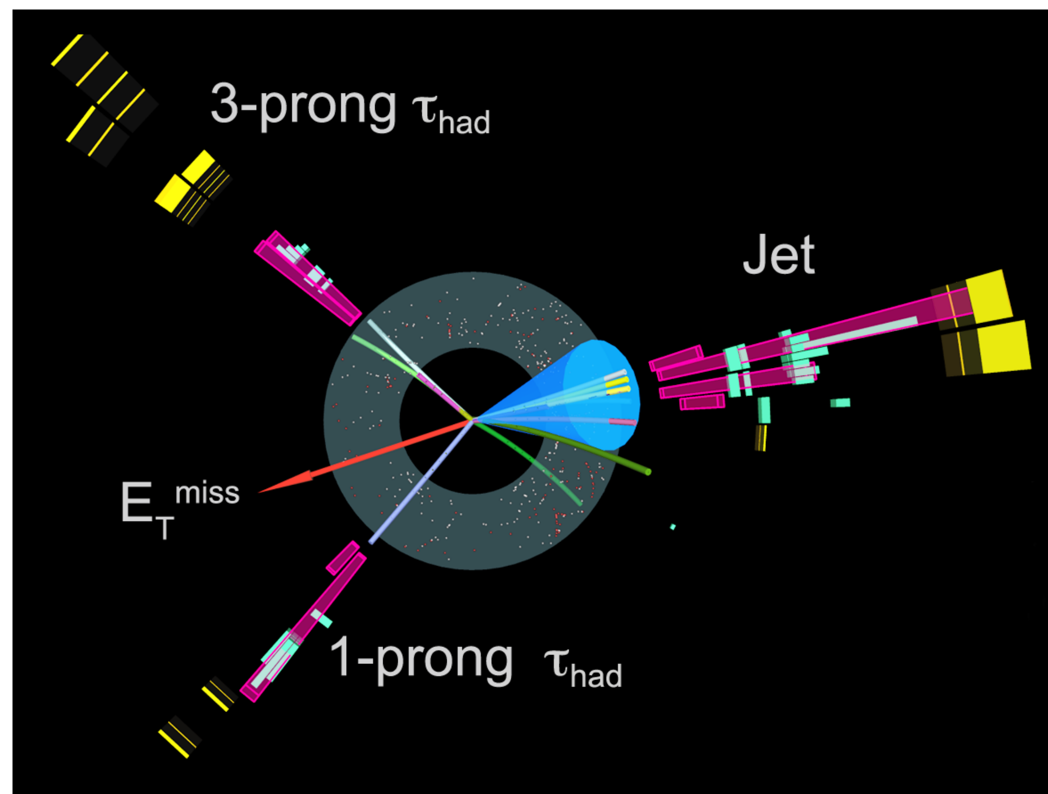
ATLAS High-mass MSSM H/A → ττ search at 13 TeV

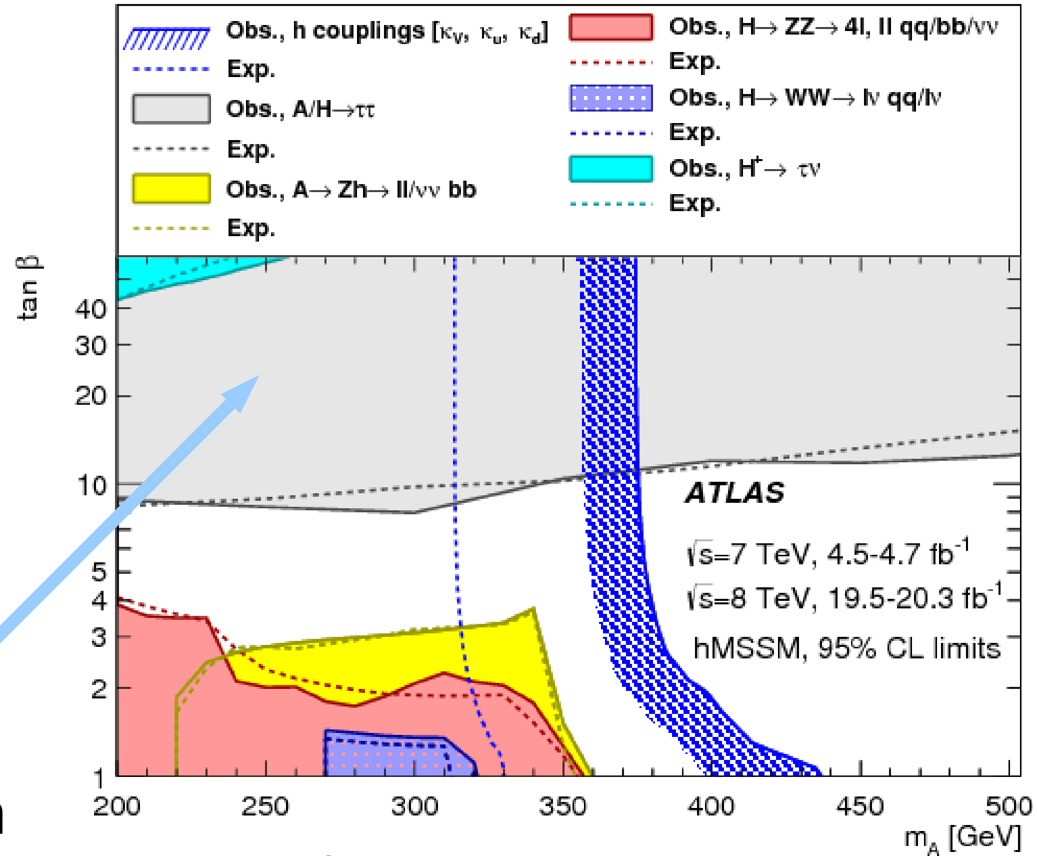
LHCP 2016

Mark Pickering
(University of Oxford)
on behalf of the
ATLAS collaboration



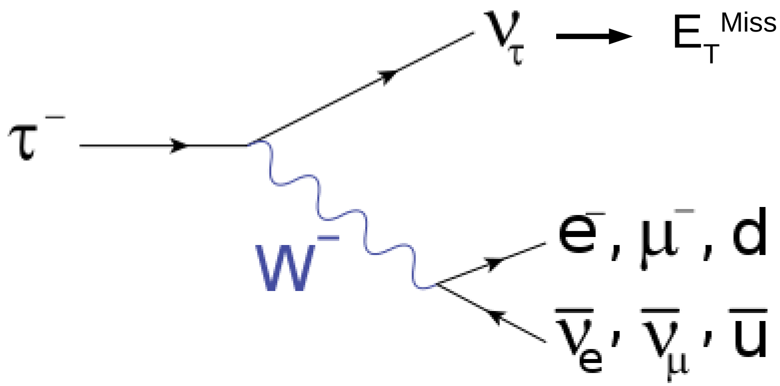
Motivation

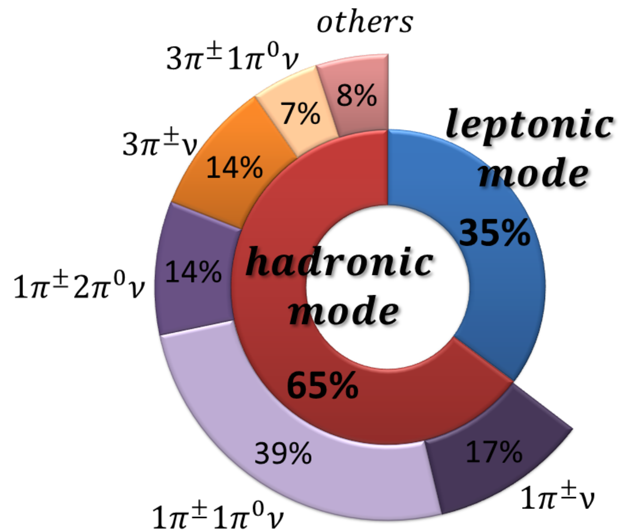
- MSSM Higgs sector:
  - two Higgs doublets
    five Higgs Bosons
  - Two charged "$H^{+/-}$", one neutral CP-odd "*A*", two neutral CP-even, "*h*" (SM-like) and "*H*"
  - Described by m_A and tanβ
    (vev ratio of doublet) at tree level
- Search focuses on the neutral H/A decaying to a pair of τ-leptons
- Channel sensitive to high tanβ region
  - tanβ > 40 for m_A ~1 TeV



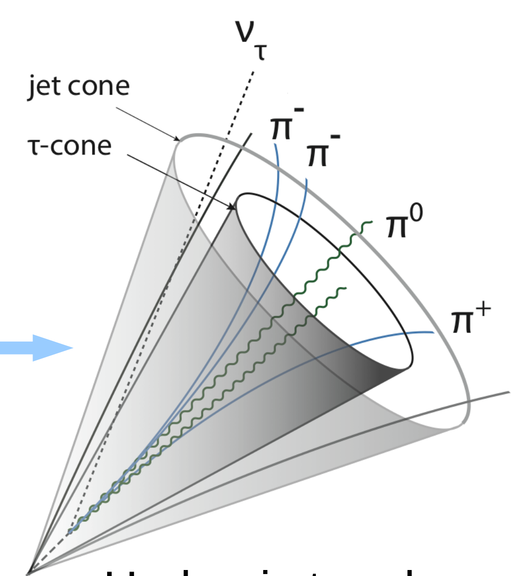
Status after Run-I
H/A → ττ analysis sensitive to unique area of parameter space (grey)

Tau Particles In ATLAS





- Decay hadronically or leptonically
- Jets form a major background
- Boosted decision tree used to identify τ_{had} and reject jets
- Discriminating variables include narrowness of tau jet, tau decay length, and number of charged tracks (1 or 3)
- Missing transverse energy associated with decay, unable to fully reconstruct parent particle mass

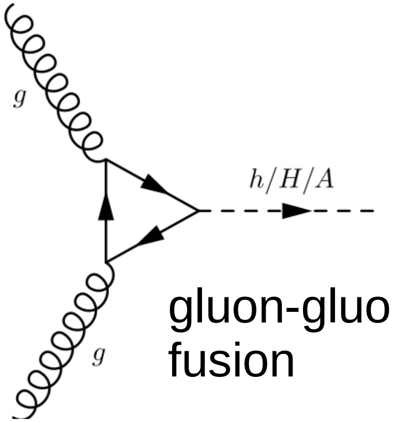


Hadronic tau decay and wider jet cone

Analysis Overview

- Analysis split by di-tau decay
  - Fully-hadronic (Had-had) and lepton+hadron (Lep-had) decay
- Search for excess of events over SM prediction
  - Successful background modelling key!
- Use a likelihood function binned in m_T^{tot} : mass discriminant (Lorenz invariant vector sum in transverse plane)

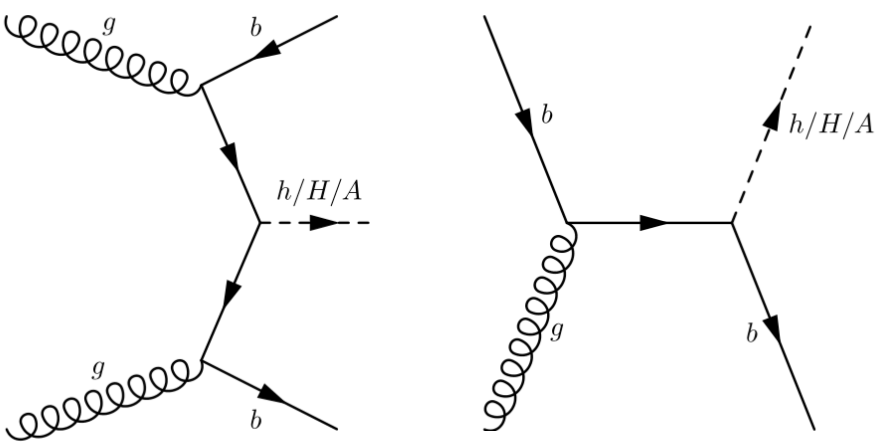
$$m_T^{tot} = \sqrt{m_T^2(\tau_1,\tau_2) + m_T^2\left(\tau_1, E_T^{Miss}\right) + m_T^2\left(\tau_2, E_T^{Miss}\right)}$$



gluon-gluon fusion

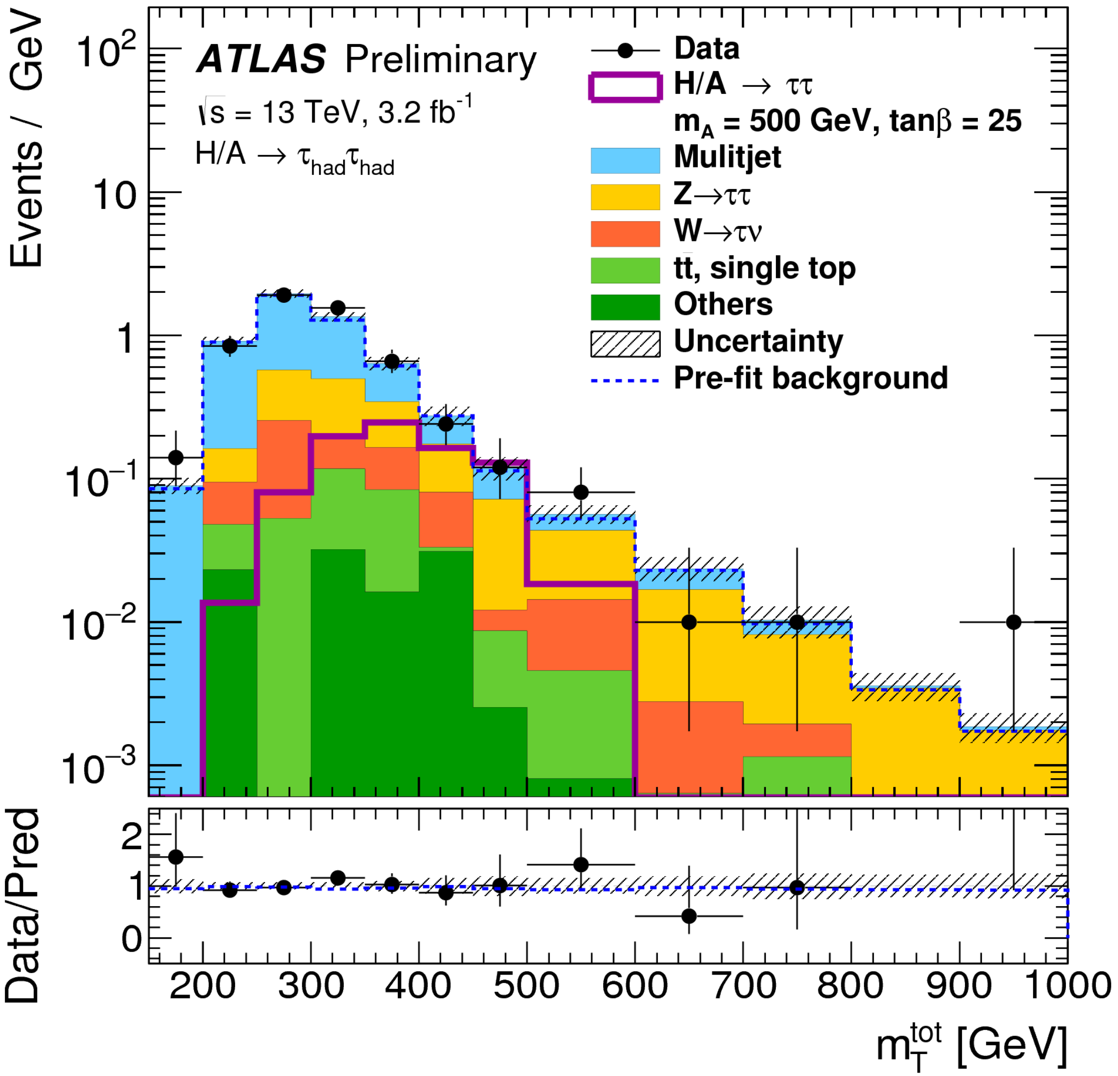
- Simulated signal with Higgs mass range 200 GeV to 1.2 TeV
  - Gluon-gluon fusion and *b*-associated production mechanisms considered
  - b-associated production increasingly important for high tanβ
- Interpret results in m_A-tanβ space

Example lowest order production mechanisms



b-associated production

Had-had – Signal Region



Signal region events as a function of the total transverse mass, 500 GeV signal in purple

Signal Region selection
Trigger: one τ_{had} $p_T > 125$ GeV
Veto events with e/μ
$p_T^{lead-\tau} > 135$ GeV, $p_T^{sublead-\tau} > 55$ GeV
Taus pass BDT identification
Δφ (lead-τ, sublead-τ) > 2.7
Taus opposite charge

Control Region:
Same sign charge taus

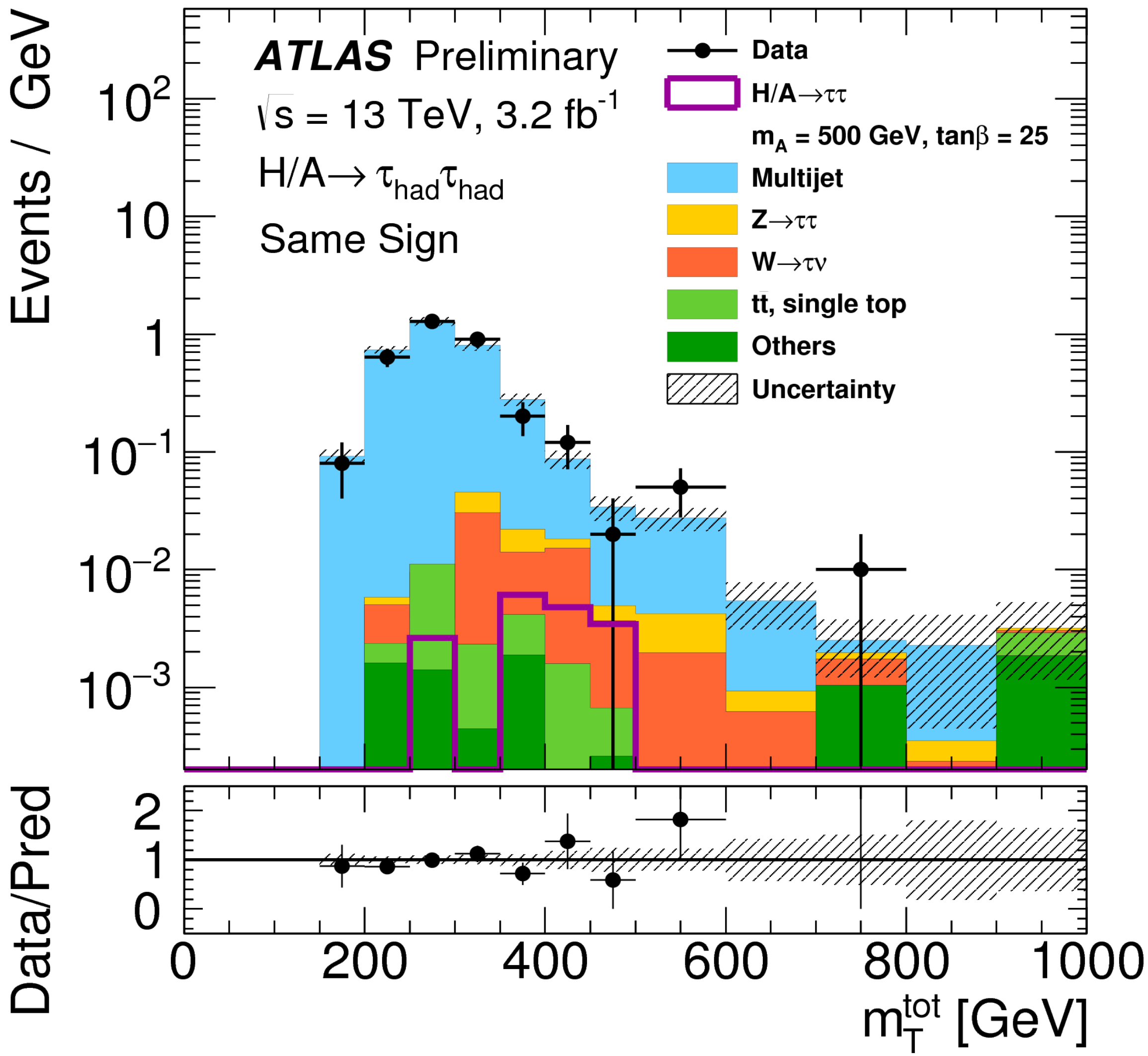
- Dominant backgrounds
  - Jet → τ_{had} fakes from multijet processes (blue) – data-driven
  - Z → ττ (yellow) – simulation
  - Non-multijet jet → τ_{had} fake sources e.g. W+jet, top (red/green) – simulation with data driven correction

Had-had – Multijet Background

- Inadequate modelling of jet → τ_{had} fakes in simulation
- Data driven multijet background (blue) estimate via "tag and probe" method in multi-jet enriched CR
  - "Tag" a jet - jet trigger
  - "probe τ" (a jet) similar to SR τ
- Calculate "fake factor" (FF), ratio of pass/fail τ-ID on "probe τ"

$$FF(p_T, N_{track}) \equiv \left.\frac{N^{pass\ \tau-ID}(p_T, N_{track})}{N^{fail\ \tau-ID}(p_T, N_{track})}\right|_{multijet}$$

- Apply FF to data events failing sublead τ-ID requirement for multijet contribution



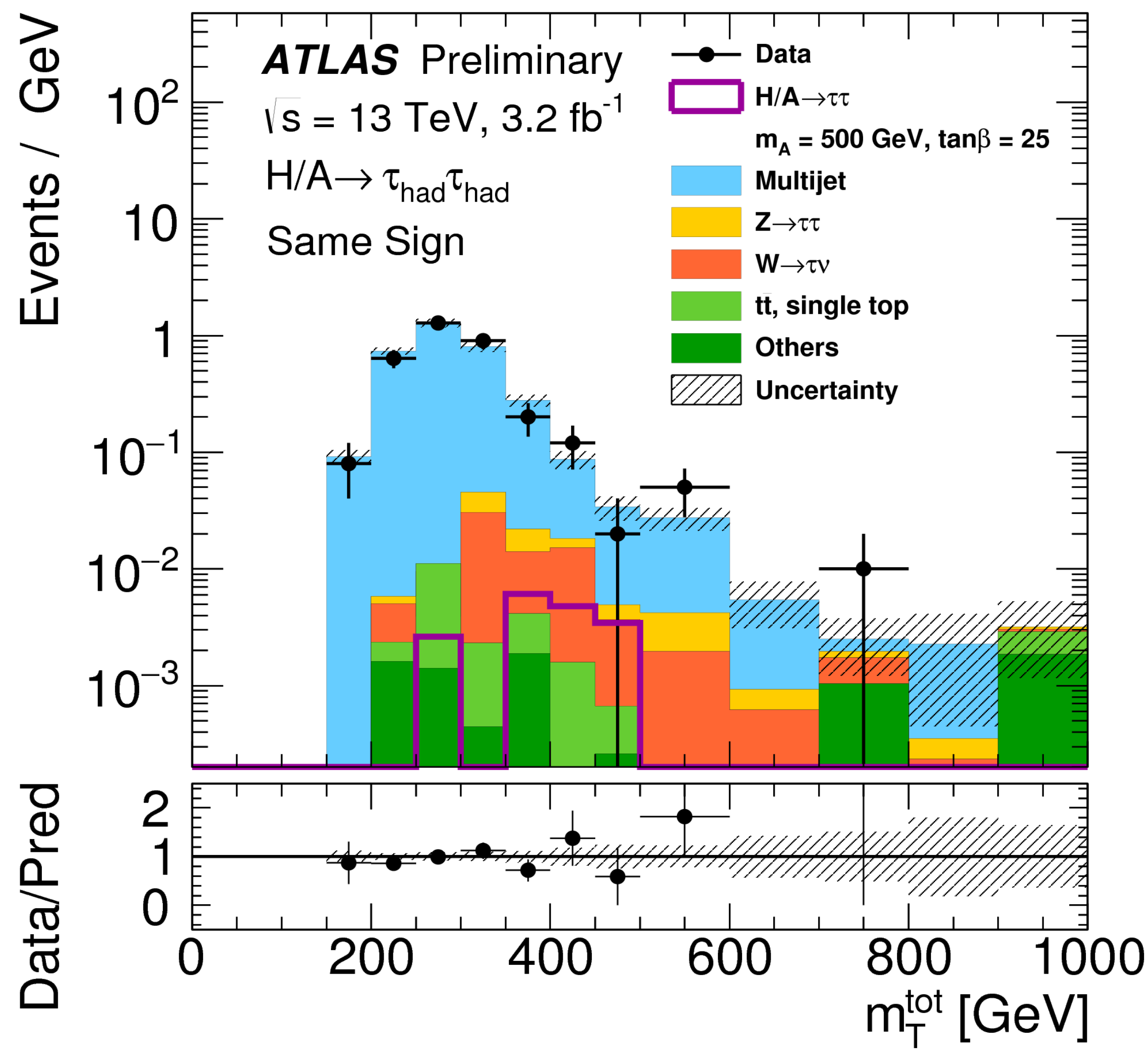
Same sign charge control region events as a function of the total transverse mass

Had-had – Non-multijet Background
Jet → τ Fakes

- Inadequate modelling of jet → τ_{had} fakes in simulation
- Data driven correction to simulation for non-multijet backgrounds via "tag and probe" method in W(→μν)+jets dominated CR
  - "Tag" a muon - muon trigger
  - "probe τ" (a jet) similar to SR τ
- Calculate "Fake rate" (FR), % jets passing τ-ID on "probe τ"

$$FR(p_T, N_{track}) \equiv \left. \frac{N^{pass\ \tau-ID}(p_T, N_{track})}{N^{all\ \tau-ID}(p_T, N_{track})} \right|_{W(\to\mu\nu)+jets}$$

- Apply fake-rate in place of BDT ID to non-truth matched taus (largest source W+jets – red/dark green)



Same sign charge control region events as a function of the total transverse mass

Lep-had – Signal Region

Signal Region Selection

Object Selection

Electron or muon trigger

\# e/μ = 1, p_T^{lep} >30 GeV, isolated

\# taus ≥ 1, p_T^{τ} >20 GeV

Tau passes BDT Identification

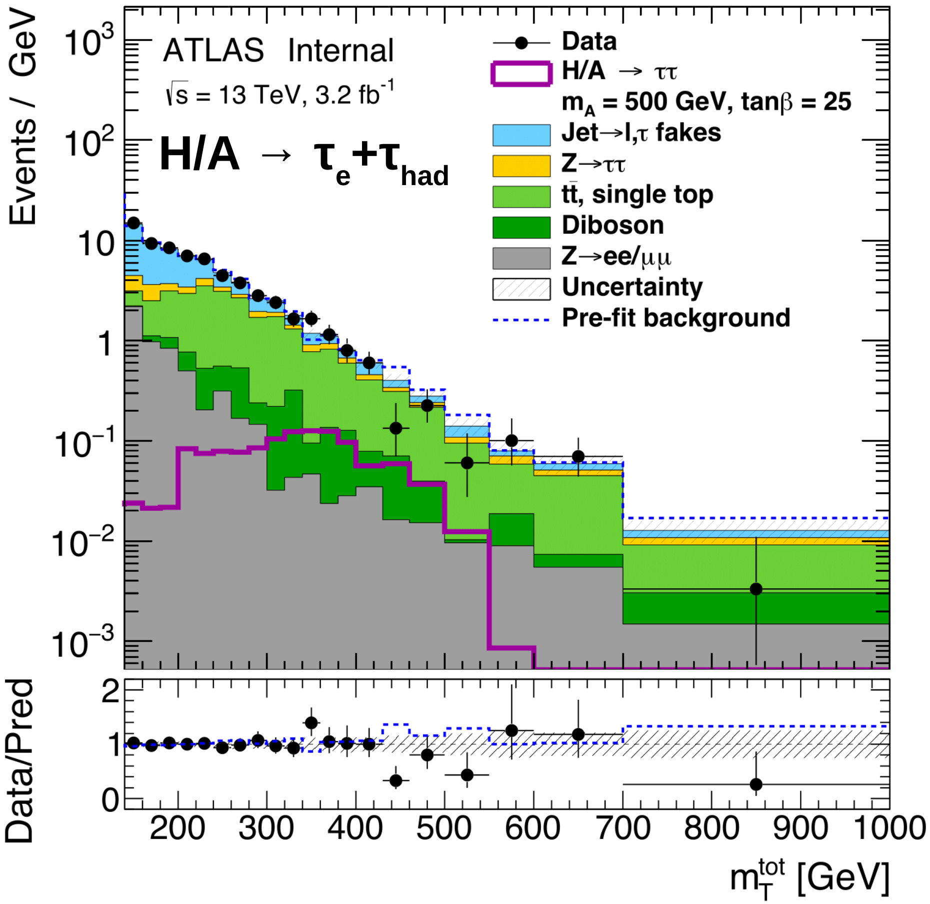
ΔΦ (lep,τ)> 2.4

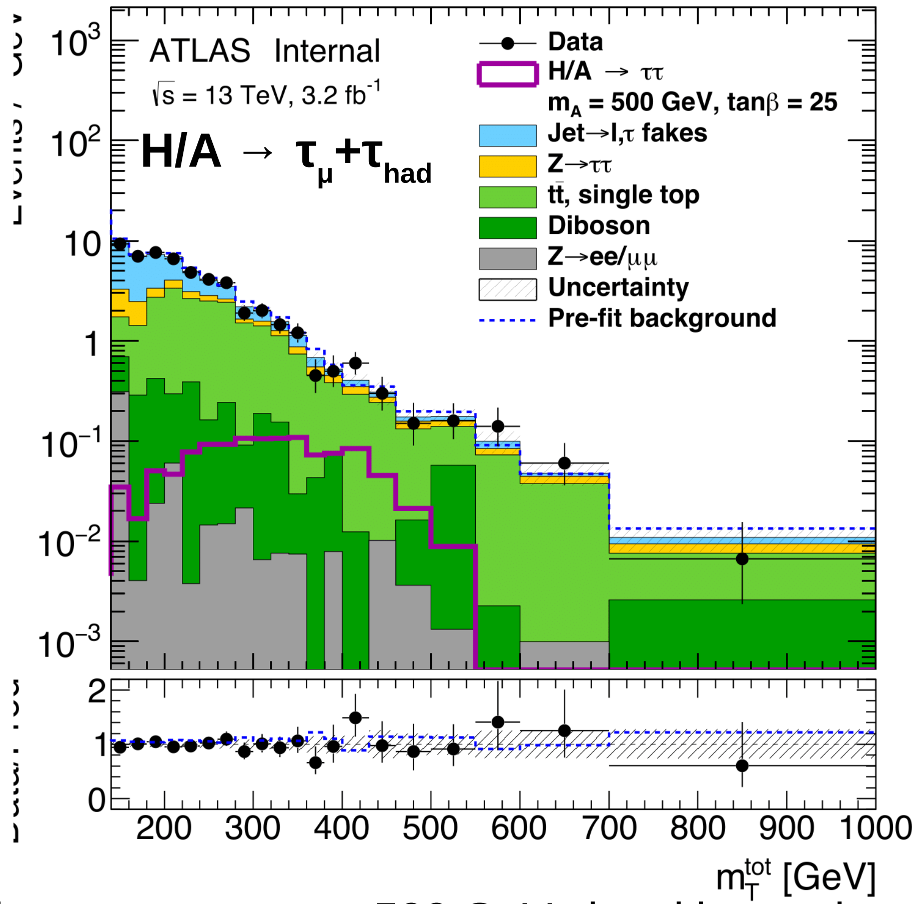
Transverse mass window :

$m_T(lep, E_T^{MISS})$ <40 GeV OR > 150 GeV

e + τ_{had} : m_{vis} exclusion window

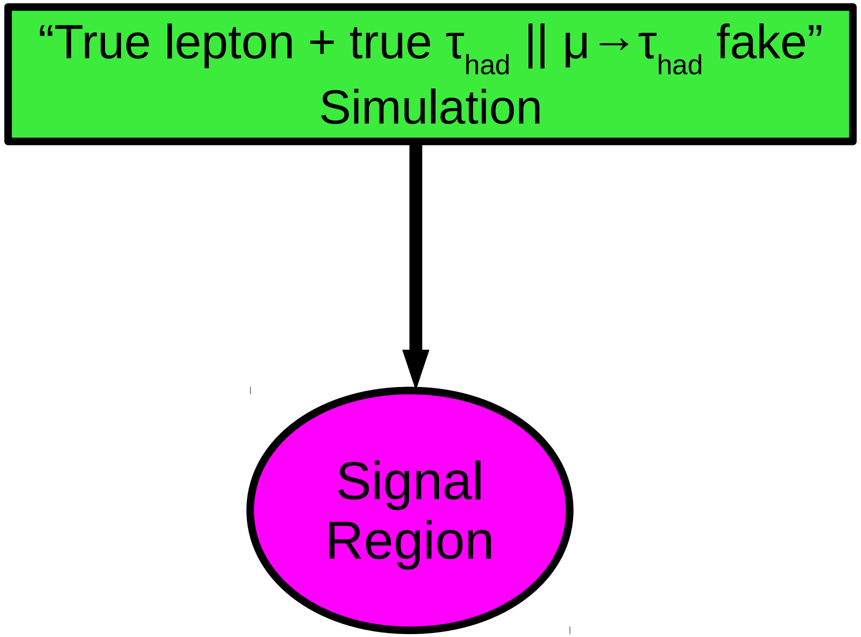
around Z peak (80-110 GeV)



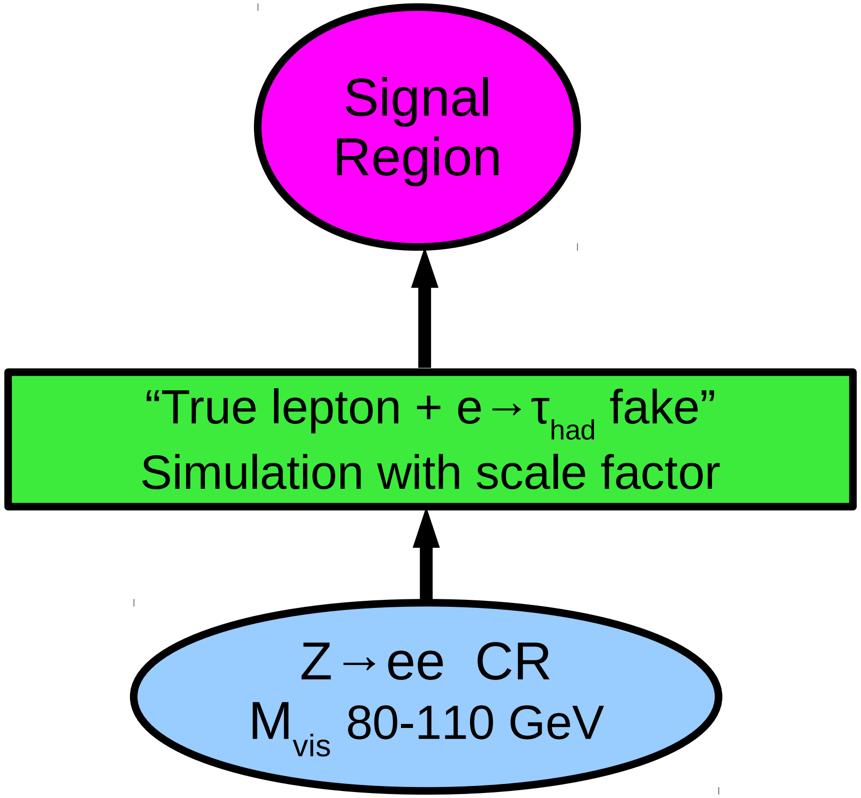


Signal region events as a function of the total transverse mass, 500 GeV signal in purple

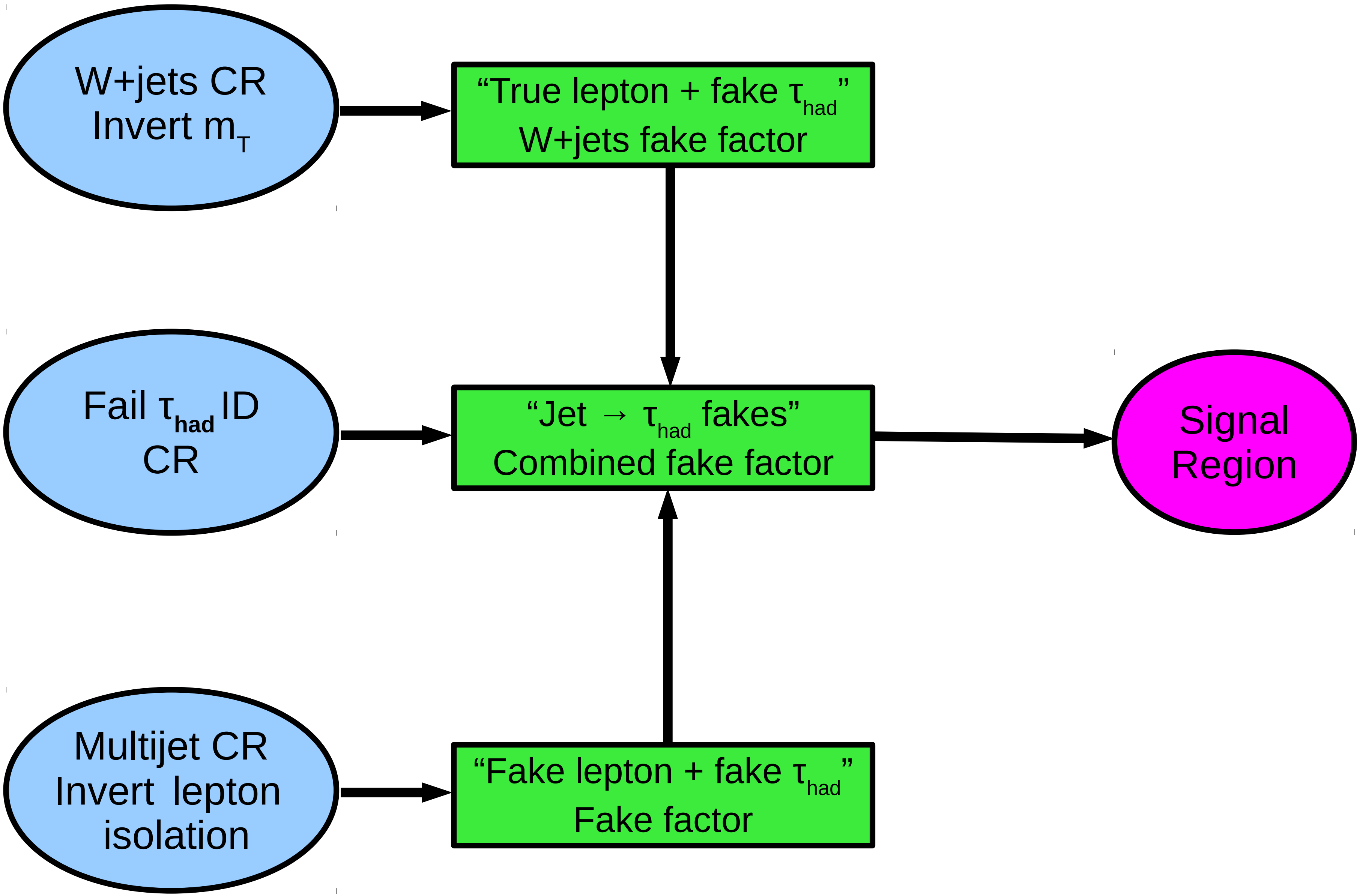
Lep-had – Background Overview



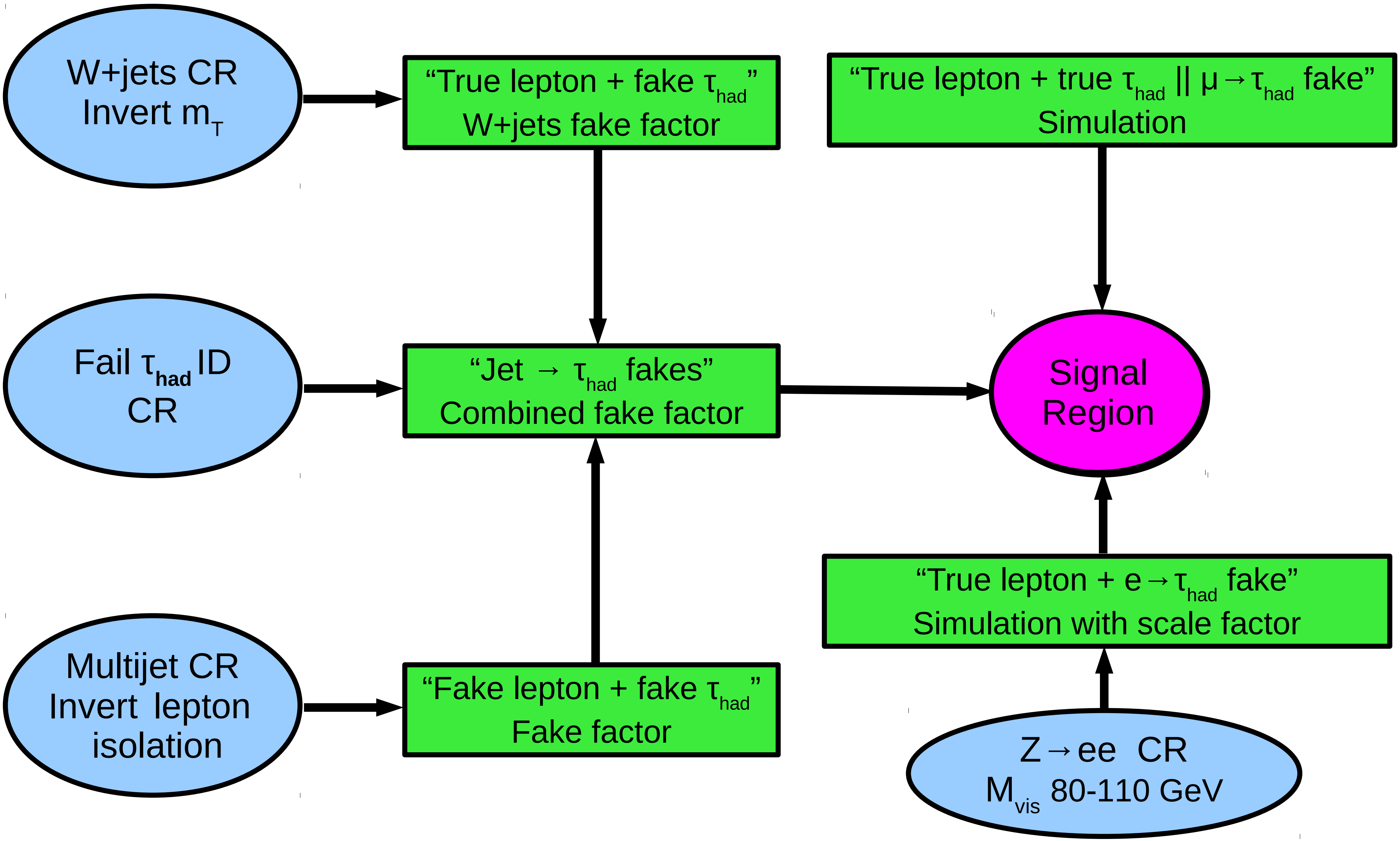
Lep-had – Background Overview



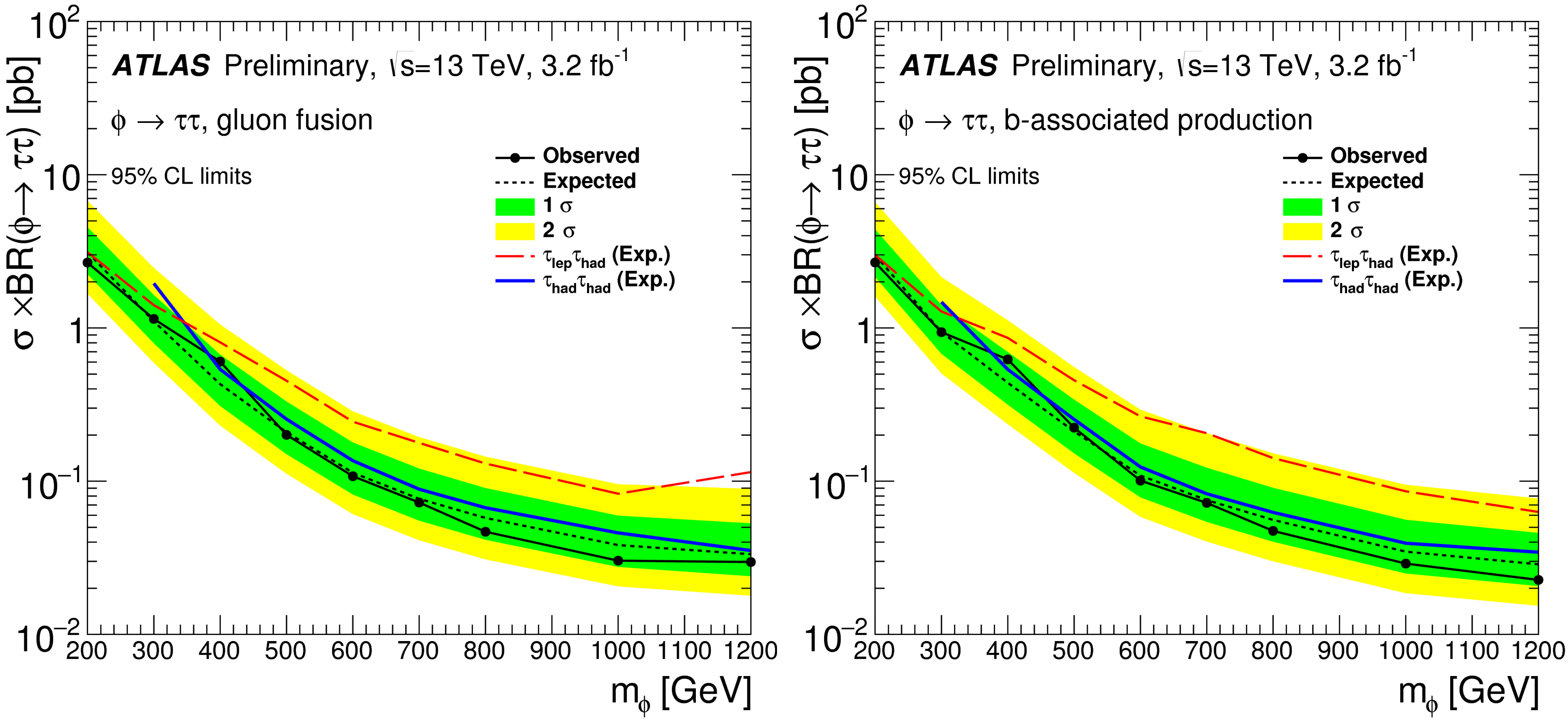
Lep-had – Background Overview



Lep-had – Background Overview

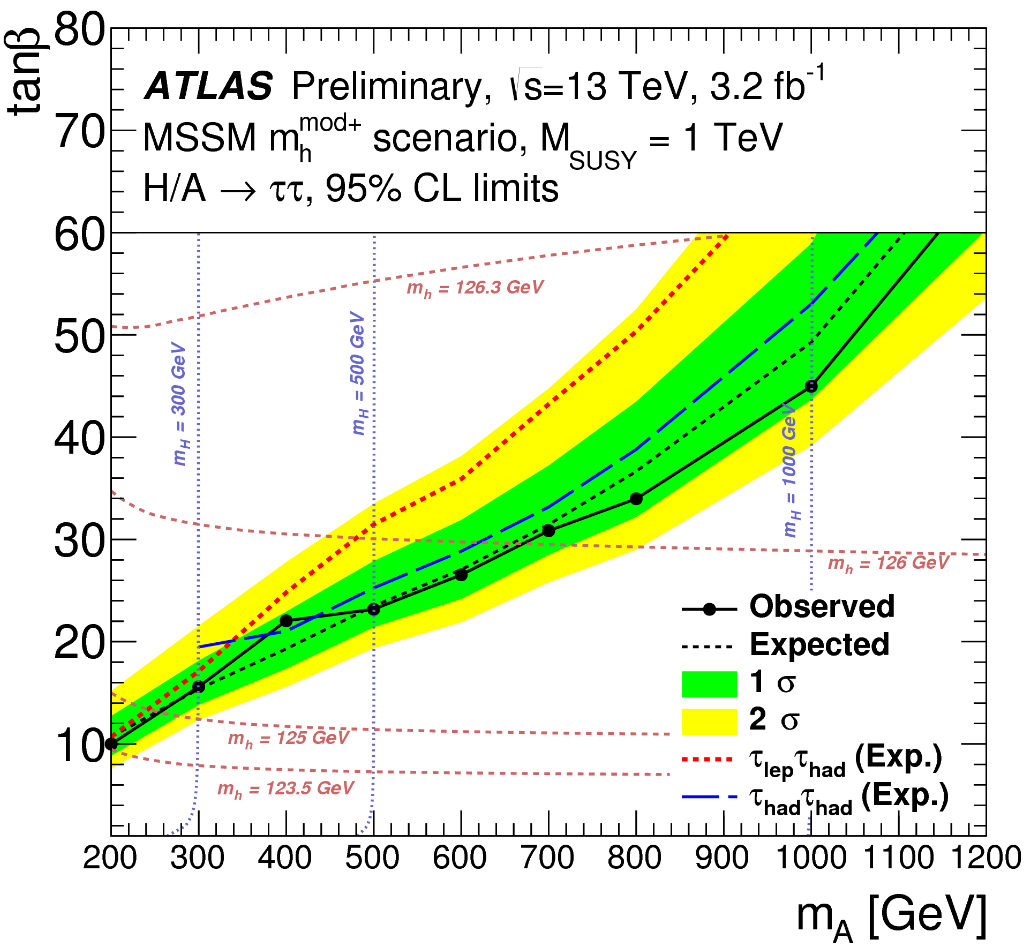


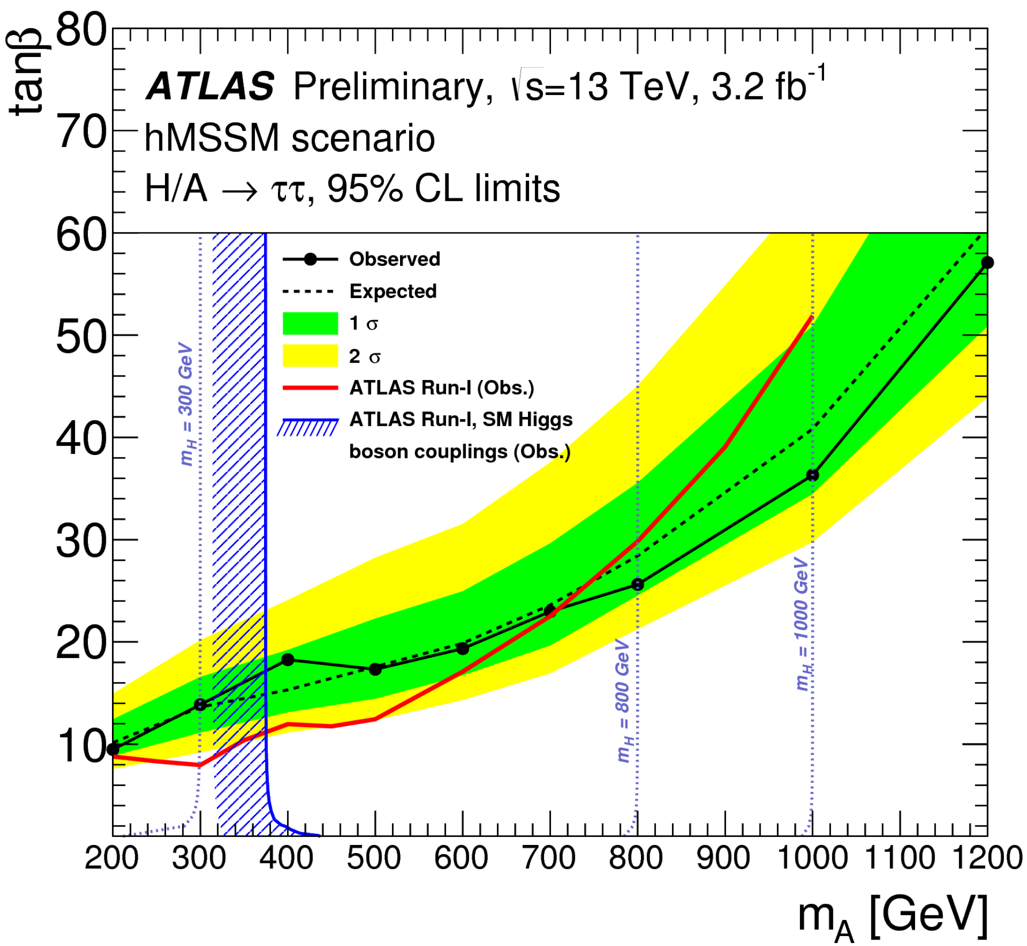
Limit on cross-section x BR



- Had-had (blue) drives sensitivity at high mass
  Lep-had (red) more important at low mass

Limit on cross-section x BR

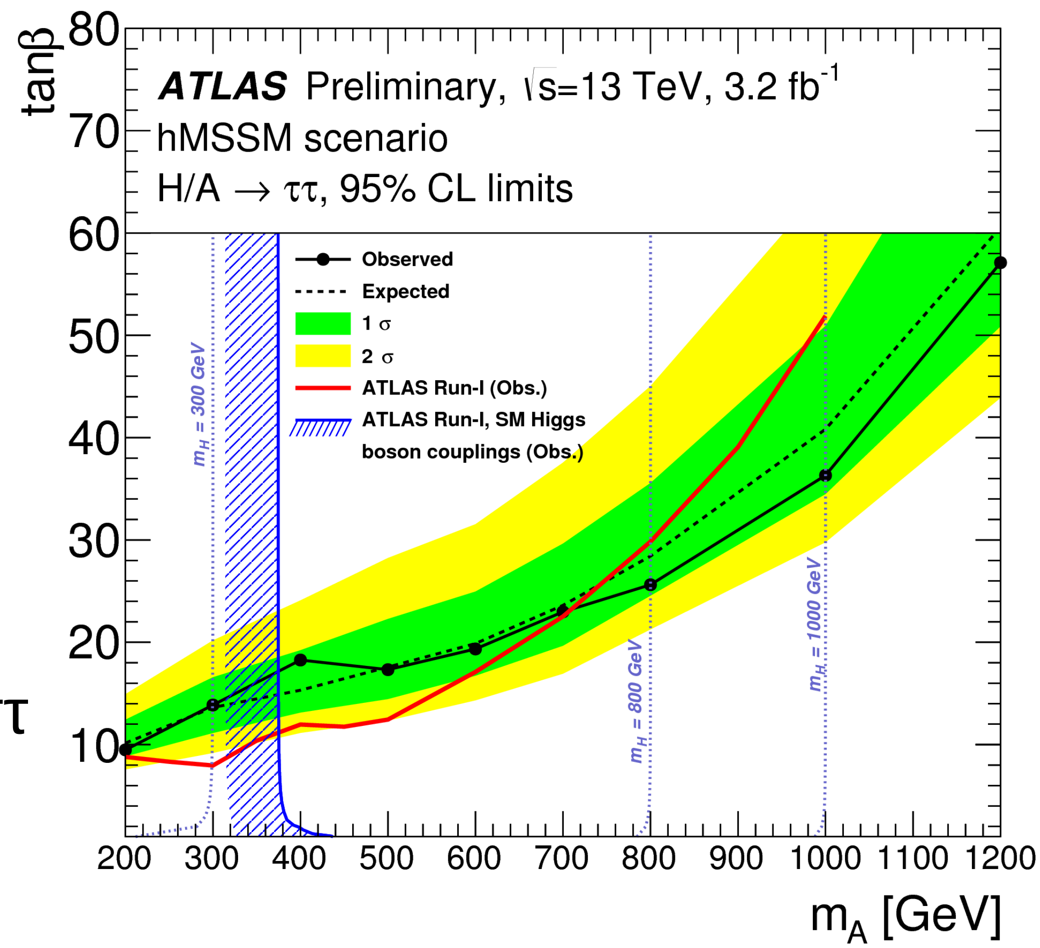




- Run-1 limit shown in red (right)
  Improvement above 700 GeV – mass reach extended to 1.2 TeV

Summary

- No significant excesses found
  - Exclusion limits placed on cross-section x BR and tanβ-m_A plane
  - ATLAS-CONF-2015-061
- Improved sensitivity compared to Run-1 above mA = 700 GeV
- 2016 an exciting year for H/A→ττ
  - Large improvement in sensitivity expected
  - Watch this space!



Backup

Systematics Overview

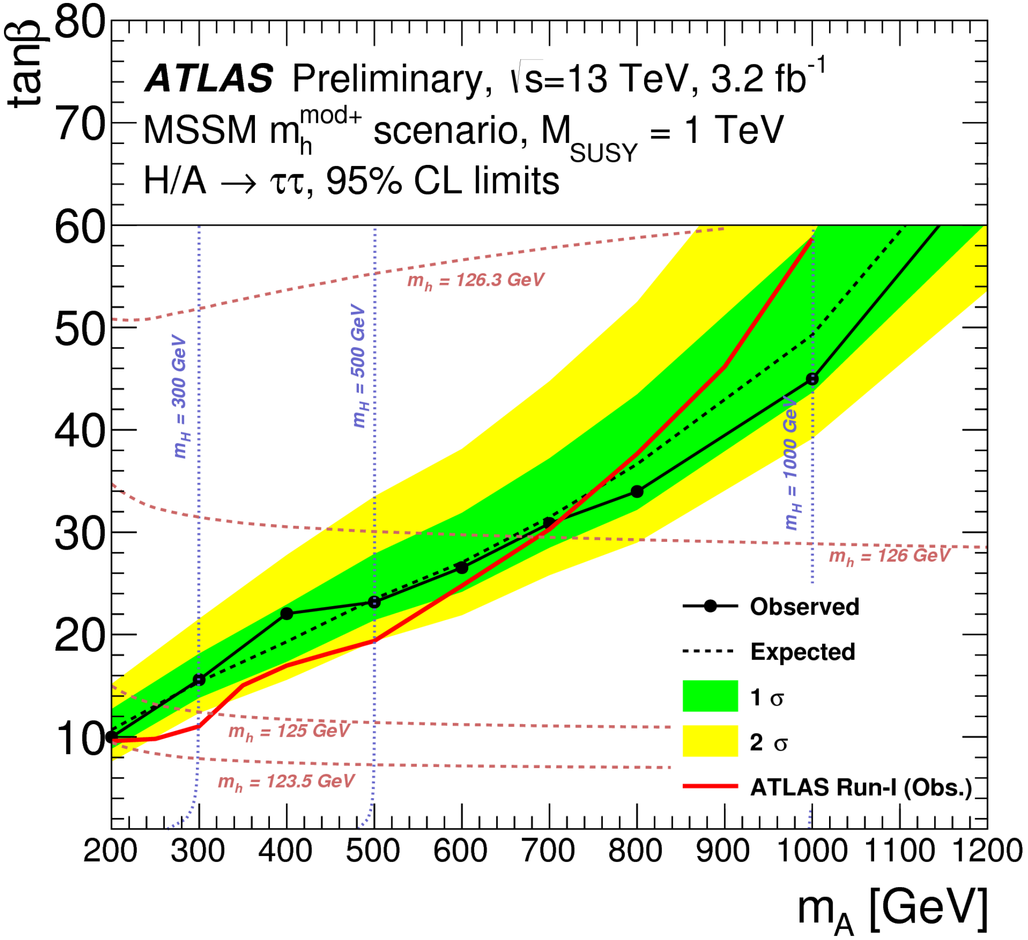
Had-had – $\tau_{had}\tau_{had}$

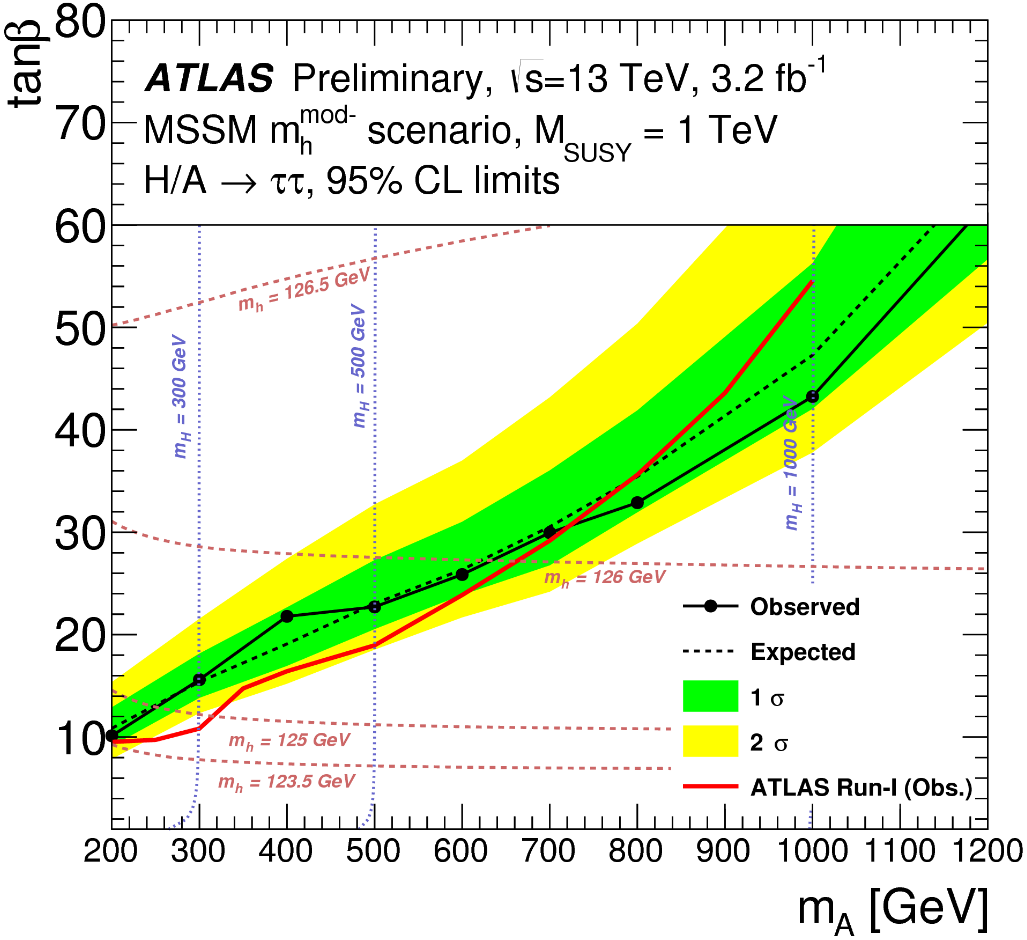
- **Multijet FF**: statistics in dijet CR + uncertainty from OS/SS difference → 7%
- **MC backgrounds with mis-ID τ_{had}**: FR uncertainty given by statistics in W(→μν)+jets CR → 9% for W(→τν)+jets
- MC-estimated samples **detector-related syst**:
  - **trigger SF**: <30%, low stats in SF measurement
  - **τ-ID, e-veto, tau reconstruction, tau energy scale, high-p_T systematics** also significant (up to 15%)

Lep-had – $\tau_{lep}\tau_{had}$

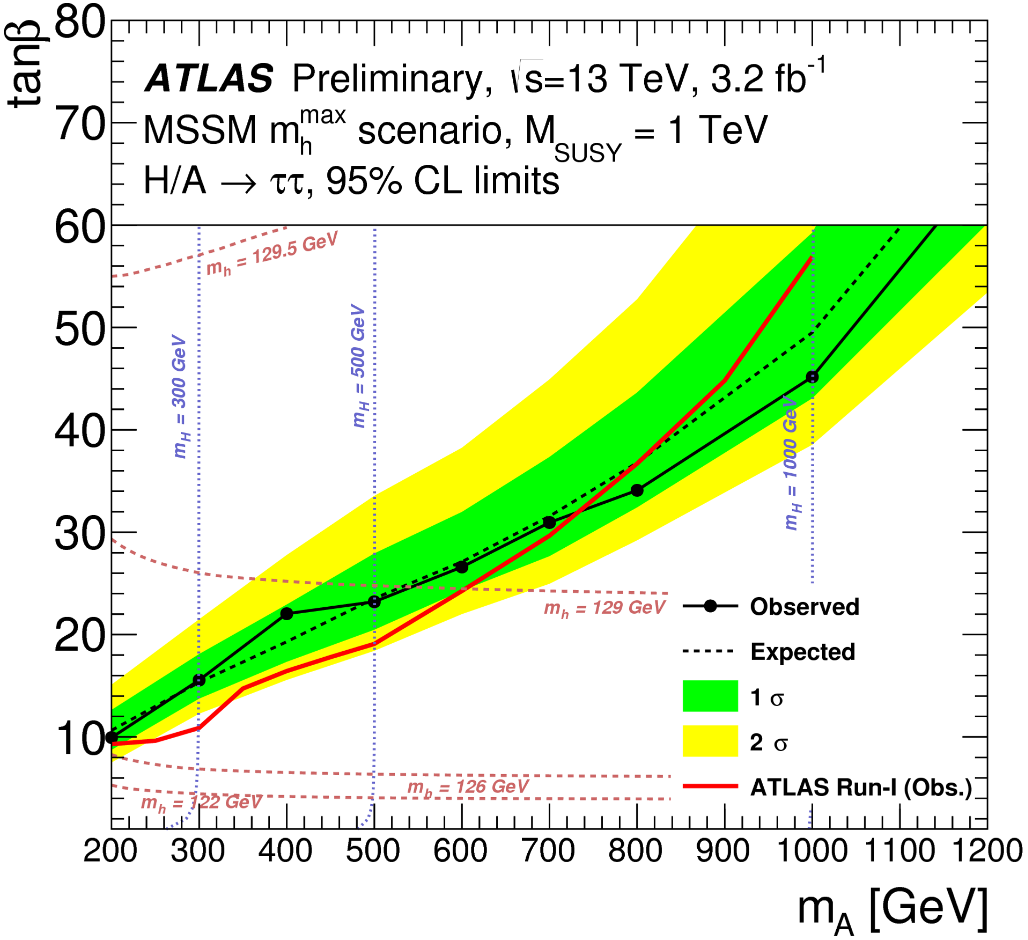
- **Multijet FF**: Stats in CRs dominate effect. True Lepton contamination in CR also contributes
- **W+jets FF**: q/g fraction between SR-CR source of syst. Also CR contamination from multijet. Total → 4-8% 1P, 5-30% 3P
- MC-estimated samples **detector-related syst**:
  - **Tau-ID, reco, e-veto, tau energy scale** significant impact on Z→ττ, top, signal (~10%) **high p_T tau systematic** also significant
  - e/μ : **trigger**, **reco, isolation, identification, energy scale** (2-5%)
  - **JES/JER** <4%, **MET syst** < 4%

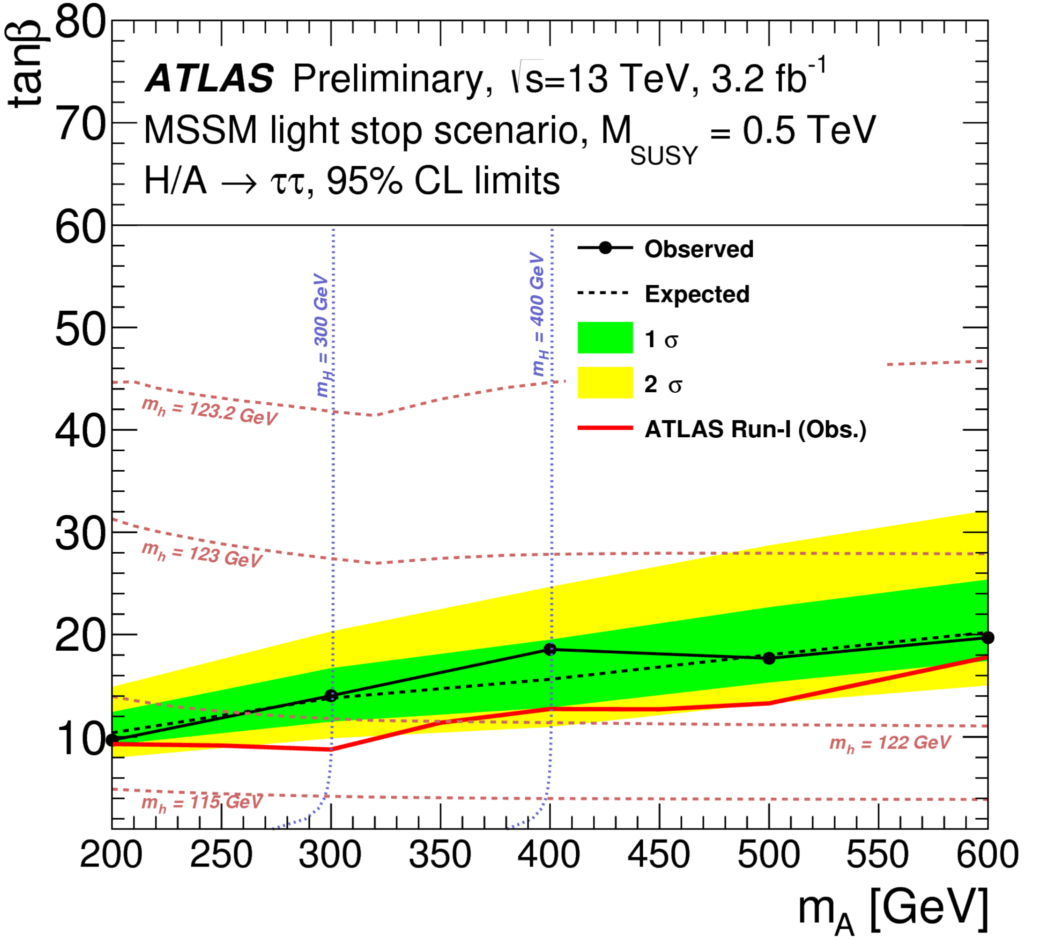
Final combined limits continued



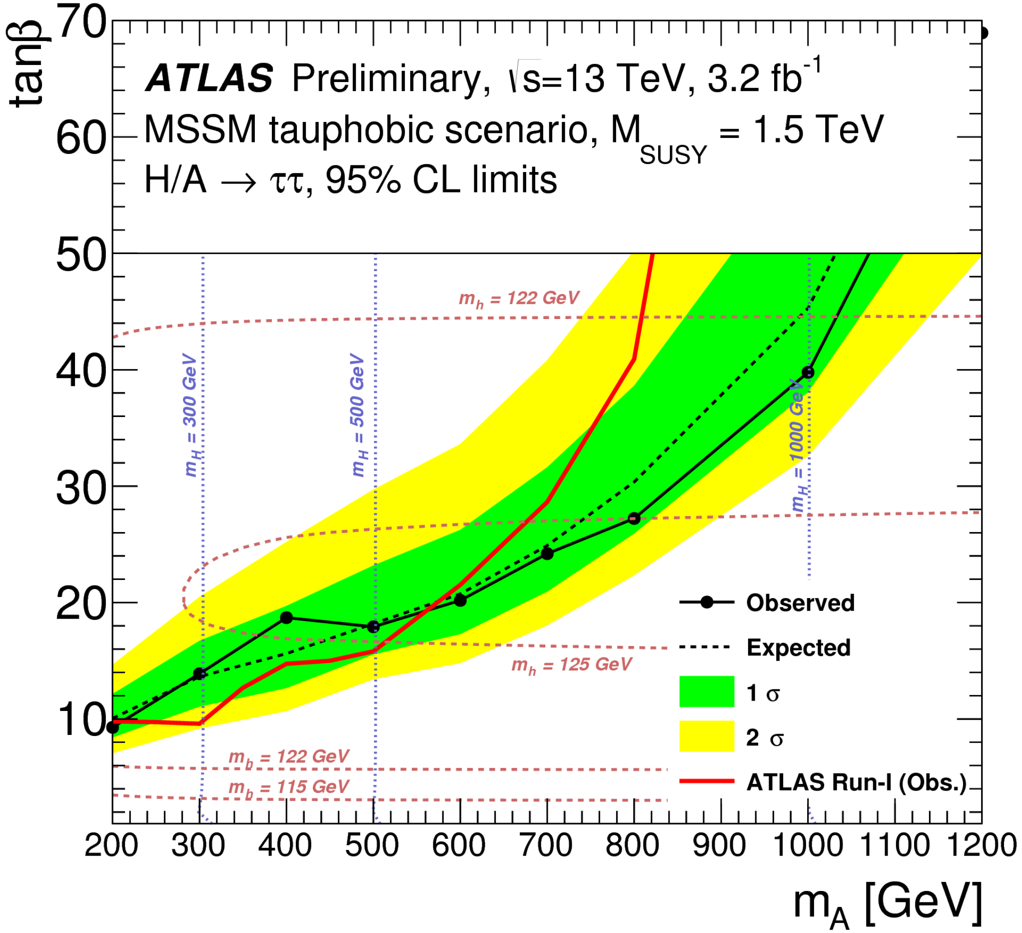


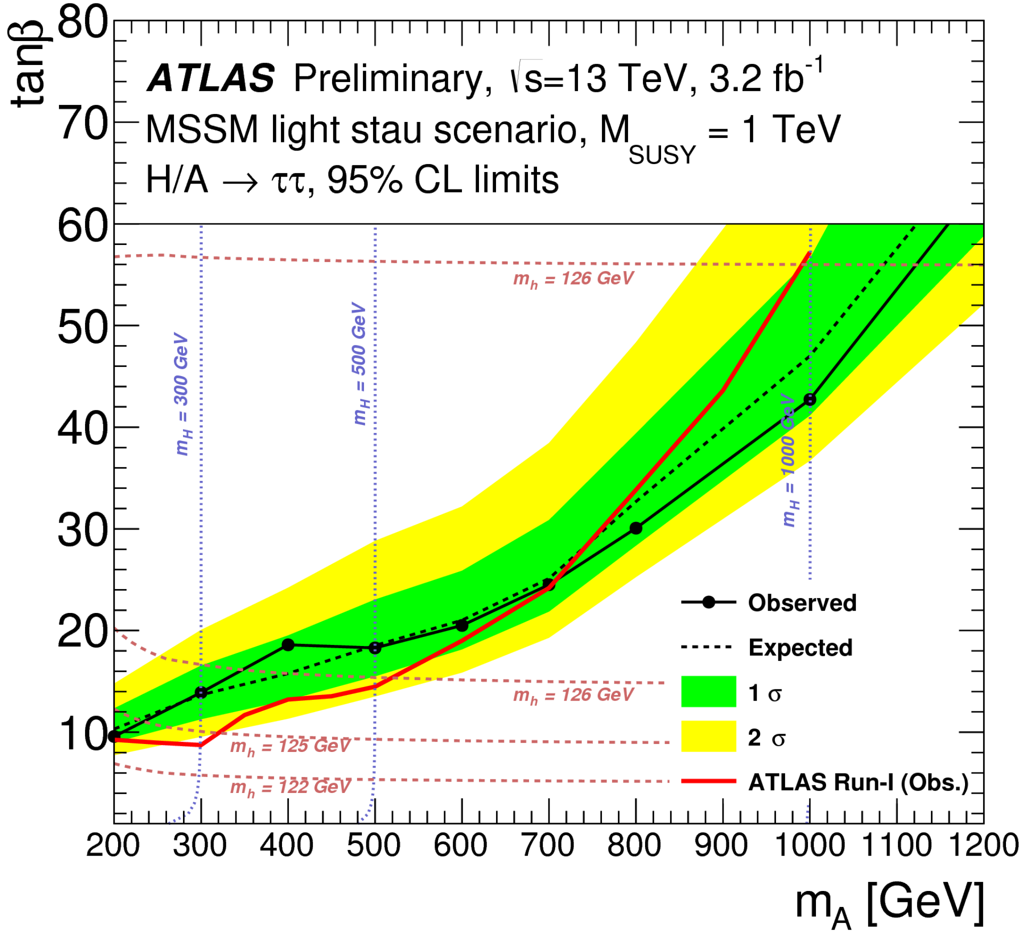
Final combined limits continued





Final combined limits continued





Final combined limits continued

m_A [GeV]	observed	expected	$+2\sigma$ [pb]	$+1\sigma$	-1σ	-2σ
200	10	11	15	13	9.1	7.9
300	16	15	22	18	13	12
400	22	19	28	23	17	15
500	23	23	33	28	21	19
600	27	27	38	32	24	22
700	31	31	45	37	28	25
800	34	37	52	43	33	29
1000	45	49	> 60	59	43	39
1200	> 60	> 60	> 60	> 60	60	53

Final number of events

Table 1: Observed number of events and background predictions after the full selections for the $\tau_e\tau_{\text{had}}$, $\tau_\mu\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. The combined statistical and systematic uncertainties are quoted.

$\tau_e\tau_{\text{had}}$ Channel	Yield
$Z \to \tau\tau$+jets	5650 ± 750
Fake τ_{had}	9640 ± 490
$Z \to \ell\ell$+jets	1390 ± 830
Top	543 ± 86
Diboson	102 ± 22
Total prediction	17300 ± 1300
Data	17480
$\tau_\mu\tau_{\text{had}}$ Channel	Yield
$Z \to \tau\tau$+jets	6720 ± 980
Fake τ_{had}	5840 ± 420
$Z \to \ell\ell$+jets	710 ± 92
Top	552 ± 80
Diboson	105 ± 22
Total prediction	14000 ± 1100
Data	13374
$\tau_{\text{had}}\tau_{\text{had}}$ Channel	Yield
$Z \to \tau\tau$+jets	52 ± 18
Multijet	175 ± 13
$W \to \tau\nu$+ jets	23.7 ± 9.6
Top	11.6 ± 5.0
Others	4.5 ± 2.4
Total prediction	268 ± 25
Data	284

MC samples

Had-had – $\tau_{had}\tau_{had}$ Lep-had – $\tau_{lep}\tau_{had}$

Signal samples

gluon-gluon fusion Powheg+Pythia8

bbH aMC@NLO+Pythia8 (AtlFast-II)

300 - 1200 GeV mass points

Background

Z+jets + $Z\rightarrow\tau\tau$ – Powheg+Pythia8 in boson mass slices

ttbar + top – Powheg+Pythia6

Diboson samples – Sherpa

Had-had	Lep-had
W+jets samples – Sherpa – Lead tau p_T slices – Various corrective factor for mismodelling requried	W+jets samples – Powheg+Pythia8

Had-had – Signal Region

Signal Region selection
recommended GRL applied
overlap removal order μ>e>τ>jet

Taus:
taus ≥ 2, 1 or 3 tracks, opposite (±1) charge
η < 2.5 (not crack region)
$p_T^{lead-\tau} > 135$ GeV
$p_T^{sublead-\tau} > 55$ GeV
$\Delta\Phi_{tau\_0,\, tau\_1} > 2.7$

trigger: HLT_tau125_medium1_tracktwo
matched to leading tau

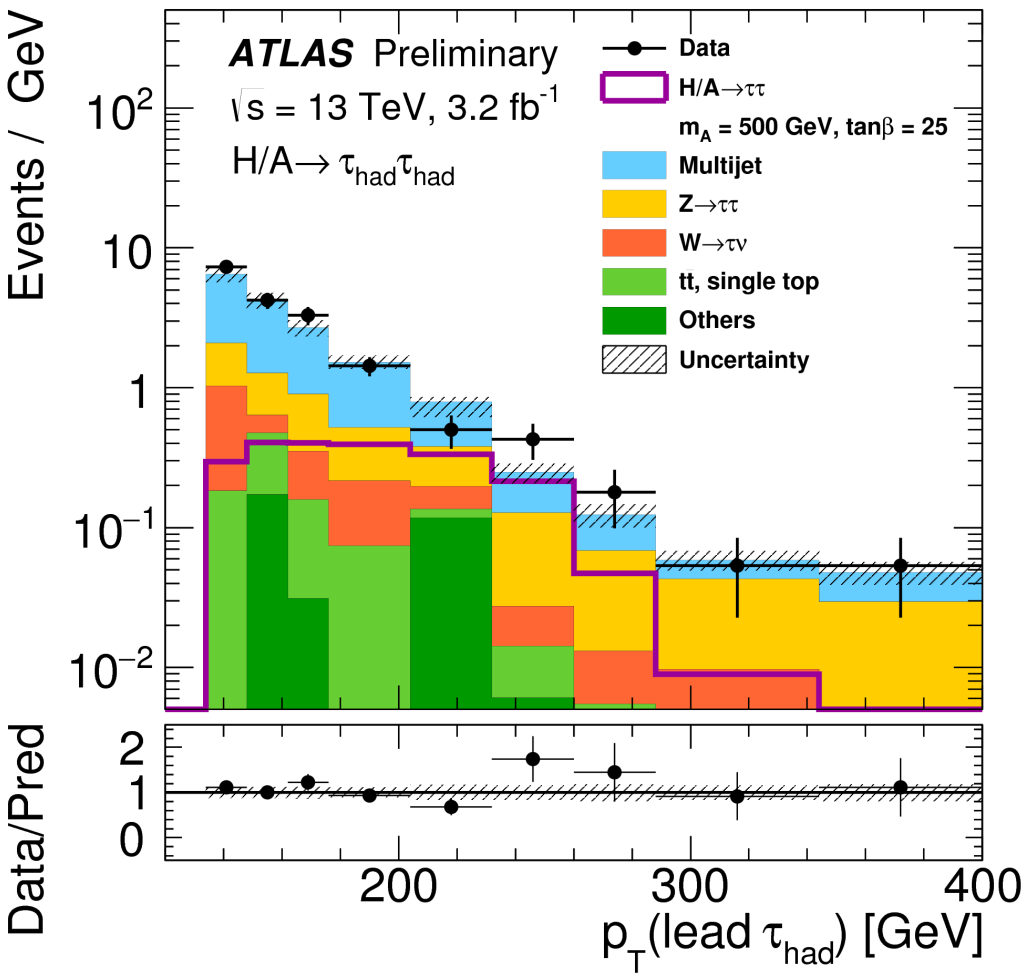
leading tau medium **isolation***

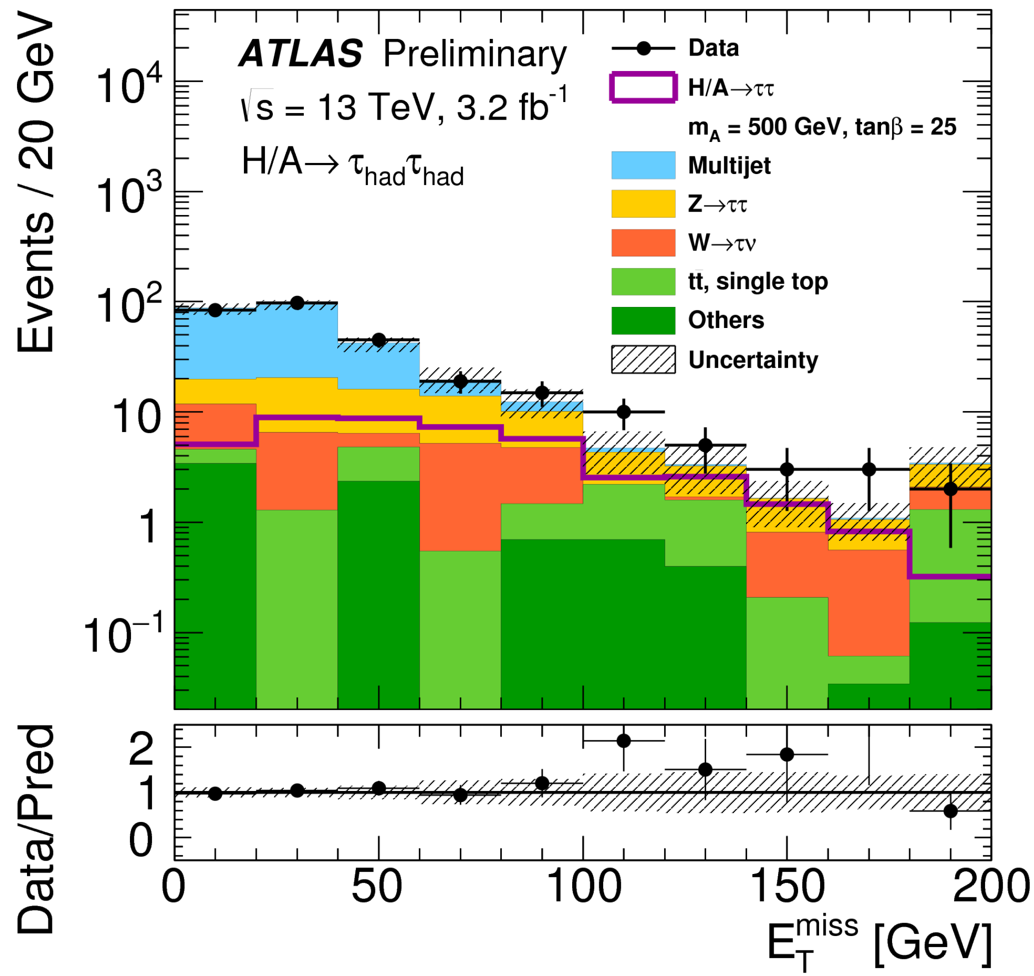
subleading pass loose **isolation***

Control Region:
same charge taus

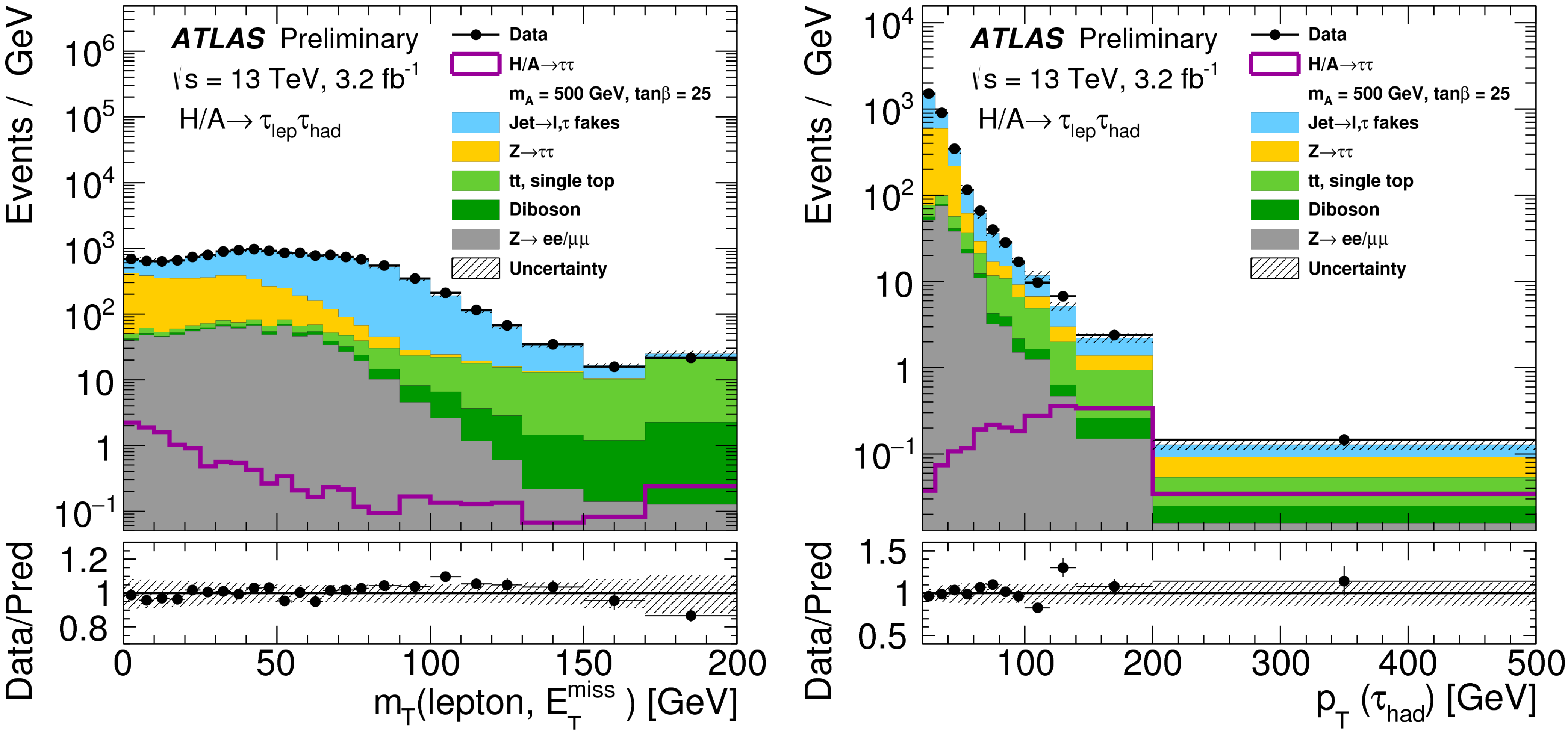
***Isolation Definition**
if data or truth-matched tau MC:
apply loose/medium τ-ID and ID SF(MC only)
non-truth matched tau MC:
apply fake rate

Had-had – Signal Region



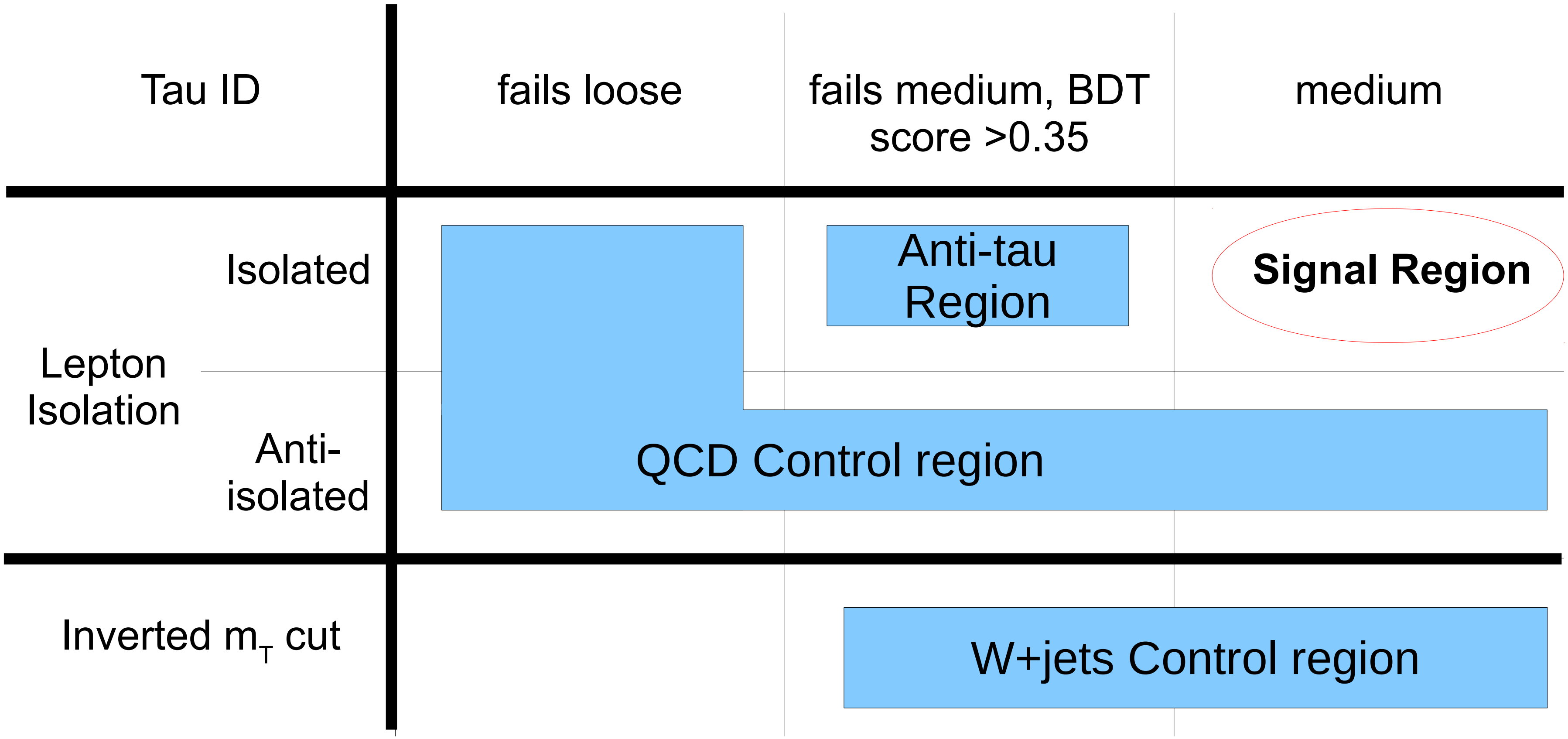


Lep-had – Signal Region



Combination
e + τ_{had} & μ + τ_{had}

Lep-had – Combined Fake Factor



$$FF(\text{comb}) = FF(W + \text{jets}) \times r_W + FF(\text{QCD}) * r_{\text{QCD}}$$

Aim: Using the known number of events in the QCD and W+jets control regions, extract the proportion of jets faking taus in the signal region from the Anti-tau region

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	medium
Lepton Isolation	Isolated	QCD Control region	Anti-tau Region	**Signal Region**
	Anti-isolated	QCD Control region	QCD Control region	QCD Control region
Inverted m_T cut			W+jets Control region	W+jets Control region

$$FF(\text{comb}) = FF(W + \text{jets}) \times r_W + FF(\text{QCD}) * r_{\text{QCD}}$$

1) Get the ratio jets faking taus in W+jets CR region - FF(W+jets)
2) Get the ratio jets faking taus in QCD CR region - FF(QCD)
3) Get the proportion of QCD jets faking taus in the anti-tau region - r_{QCD}

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	medium
Lepton Isolation	Isolated			**Signal Region**
	Anti-isolated			
Inverted m_T cut			FF_{W+jets} denominator	FF_{W+jets} numerator

$$FF(\text{comb}) = FF(W + \text{jets}) \times \quad +$$

Factor of jets in W+jets faking taus

$$FF(W + \text{jets}) = \frac{N(\text{pass “medium” tau ID})}{N(\text{fail “medium” tau ID})},$$

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	medium
Lepton Isolation	Isolated			**Signal Region**
	Anti-isolated		FF_{QCD} denominator	FF_{QCD} numerator
Inverted m_T cut				

$$FF(\text{comb}) = \qquad\qquad + FF(\text{QCD}) *$$

Factor of QCD faking taus

$$FF(\text{QCD}) = \frac{N(\text{pass “medium” } \tau \text{ identification})}{N(\text{fail “medium” } \tau \text{ identification and jet BDT} > 0.35)},$$

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	medium
Lepton Isolation	Isolated	FF_{lep} numerator		**Signal Region**
	Anti-isolated	FF_{lep} denominator		
Inverted m_T cut				

$$FF(\text{comb}) =$$

$$FF = \frac{N(\text{pass “gradient” lepton isolation})}{N(\text{fail “gradient” lepton isolation})},$$

Factor of QCD faking leptons

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	medium
Lepton Isolation	Isolated		Extract r_{QCD}	**Signal Region**
	Anti-isolated		Apply FF_{lep}	
Inverted m_T cut				

$$FF(\text{comb}) = \qquad \times r_W + \qquad * r_{QCD}$$

$$FF = \frac{N(\text{pass "gradient" lepton isolation})}{N(\text{fail "gradient" lepton isolation})},$$

Factor of QCD faking leptons

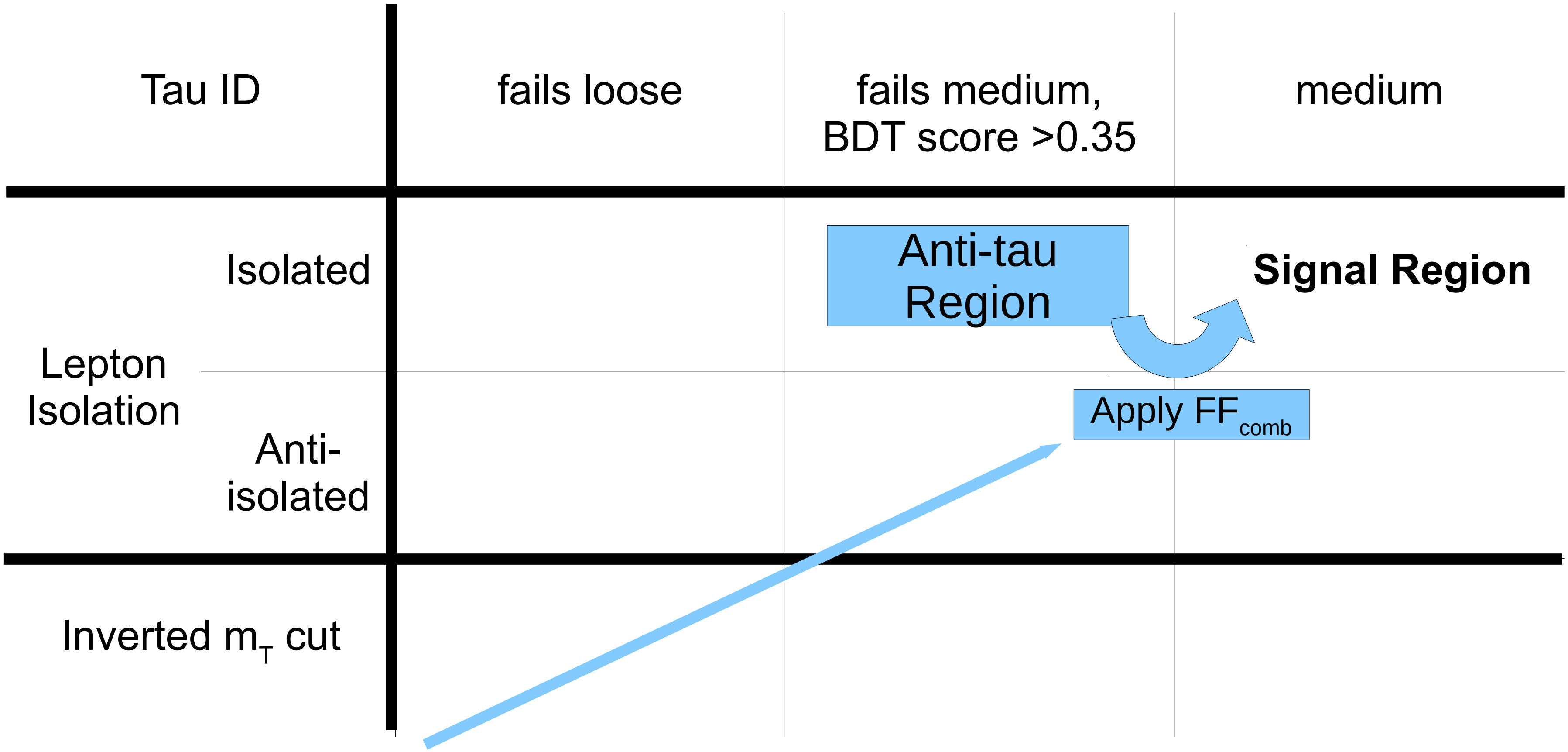
$$r_{QCD} = \frac{N(\text{multi-jet, data-driven})}{N(\text{data}) - N(\text{true } \tau_{had}, \text{MC})},$$

Proportion of QCD in Anti-tau region

$$r_W = 1 - r_{QCD}$$

Proportion of Wjets in Anti-tau

Lep-had – Combined Fake Factor



$$FF(\text{comb}) = FF(W + \text{jets}) \times r_W + FF(\text{QCD}) * r_{\text{QCD}}$$

$$FF = \frac{N(\text{pass ``gradient'' lepton isolation})}{N(\text{fail ``gradient'' lepton isolation})},$$

$$r_{\text{QCD}} = \frac{N(\text{multi-jet, data-driven})}{N(\text{data}) - N(\text{true } \tau_{\text{had}}, \text{MC})},$$

$$r_W. = 1 - r_{\text{QCD}}$$

$$FF(\text{QCD}) = \frac{N(\text{pass ``medium'' } \tau \text{ identification})}{N(\text{fail ``medium'' } \tau \text{ identification and jet BDT} > 0.35)},$$

$$FF(W + \text{jets}) = \frac{N(\text{pass ``medium'' tau ID})}{N(\text{fail ``medium'' tau ID})},$$

Lep-had – Combined Fake Factor

Tau ID		fails loose	fails medium, BDT score >0.35	Medium
Lepton Isolation	Isolated	FF_{lep} numerator	Extract r_{QCD}	**Signal Region**
	Anti-isolated	FF_{lep} denominator	FF_{QCD} denominator	FF_{QCD} numerator
Inverted m_T cut			FF_{W+jets} denominator	FF_{W+jets} numerator

Apply FF_{lep}

Apply FF_{comb}

$$FF(\text{comb}) = FF(W + \text{jets}) \times r_W + FF(\text{QCD}) * r_{\text{QCD}}$$

$$r_{\text{QCD}} = \frac{N(\text{multi-jet, data-driven})}{N(\text{data}) - N(\text{true } \tau_{\text{had}}, \text{MC})},$$

$$FF(\text{QCD}) = \frac{N(\text{pass ``medium'' } \tau \text{ identification})}{N(\text{fail ``medium'' } \tau \text{ identification and jet BDT} > 0.35)},$$

$$FF(W + \text{jets}) = \frac{N(\text{pass ``medium'' tau ID})}{N(\text{fail ``medium'' tau ID})},$$

$$FF = \frac{N(\text{pass ``gradient'' lepton isolation})}{N(\text{fail ``gradient'' lepton isolation})},$$

$$r_{\text{QCD}} = 1 - r_W$$

Mass Reconstruction - Algorithms

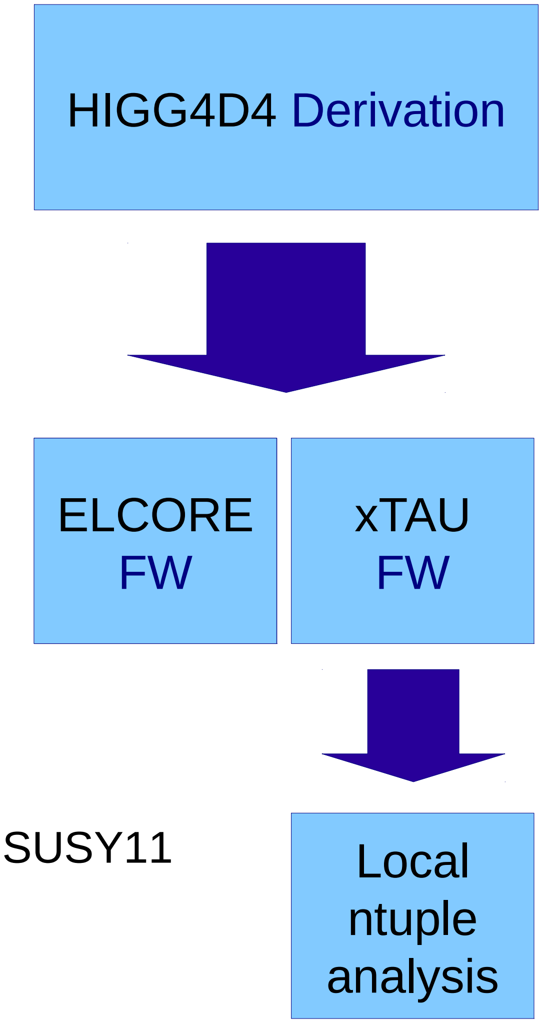
- Various mass reconstruction algorithms considered – tuned to high mass
- Missing Mass Calculator (MMC)
  - Assume missing transverse momentum is due entirely to the neutrinos
  - Scan over the angles between the neutrinos and the visible τ decay products
  - Each solution is weighted according to probability density functions that are derived from simulated τ decays
  - MAXW – solution of maximum weight point of phase space
  - MLM – highest probability mass in calculation
  - MLN3P – point with most likely neutrino decay
- The Matrix-element Oriented SAmplIng Calculator (MOSAIC)
  - similar technique to the MMC in using likelihood function and probability density function
  - uses a matrix element based maximum likelihood.
- Total Transverse Mass – M_{TOT}
  - Used in Run-I
- M_{TOT} considered to be the optimal mass reconstruction technique for maximising signal-background separation

$$m_T^{tot} = \sqrt{m_T^2(\tau_1, \tau_2) + m_T^2\left(\tau_1, E_T^{Miss}\right) + m_T^2\left(\tau_2, E_T^{Miss}\right)}$$

$$m_T = \sqrt{2 p_{T1} p_{T2} \left(1 - \cos\Delta\phi\right)}$$

Derivations and Analysis Frameworks

- Running on HIGG4D4* derivations:
  - 1 || 3 track tau candidates with:
  - $p_T^{lead-\tau} > 160$ GeV && $p_T^{sublead-\tau} > 45$ GeV OR
    $p_T^{lead-\tau} > 80$ GeV && $p_T^{sublead-\tau} > 50$ GeV && jet tau looseID$^{sublead-\tau}$
  - Significant reduction in time to produce ntuples
- Two xAOD based Frameworks:
  - "xTAU" Framework and Dresden based "ELCore" framework
  - xTAU produces ntuples from xAODs
  - ELCORE runs directly on derivations
- Plotting codes for ntuples vary between analysers
  - Cross checks and acceptance challenges in place



* For EOYE QCD estimation in di-jet control region we use the looser SUSY11

- Requires firing of single jet triggers – no tau ID requirement

* For jet → tau fake rate calculation use HIGG4D2 derivations

- Medium quality lepton ($p_T^e > 15$ GeV || $p_T^\mu > 12$ GeV)
  + hadronic tau $p_T^\tau > 18$ (1 || 3 track)