Data-driven background measurements in CMS

Matti Kortelainen
Helsinki Institute of Physics

on behalf of the CMS collaboration

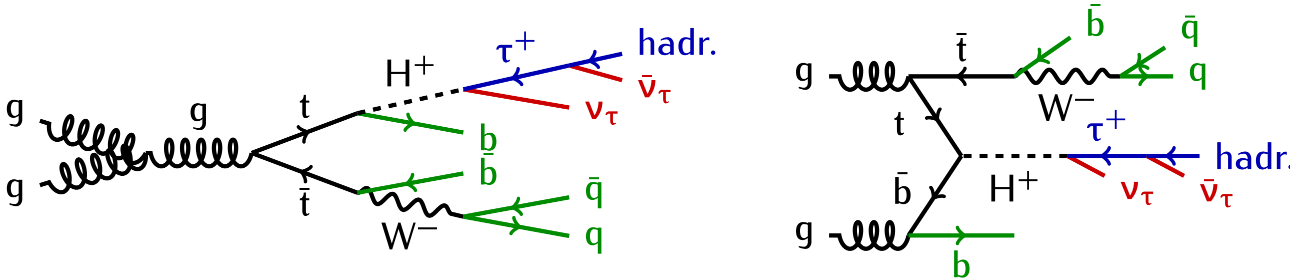
Charged2014: Prospects for Charged
Higgs Discovery at Colliders, Uppsala

September 16, 2014

Introduction

- The following backgrounds are measured from data in CMS analyses
- $H^+ \to \tau\nu_\tau$, τ_h+jets final state (Lauri's talk)
  - CMS PAS HIG-14-020
  - Multijet events (jet misidentified as τ_h)
  - EWK+$t\bar{t}$ events with genuine $\tau \to$ hadrons decay identified as τ_h
    - ⋆ In this presentation EWK+$t\bar{t}$ = SM $t\bar{t}$, $W+$jets, Z/γ^*, single top, diboson
- $H^+ \to tb$ / $H^+ \to \tau\nu_\tau$, $\mu\tau_h$ final state (Pietro's talk)
  - CMS PAS HIG-13-026
  - Jet misidentified as τ_h ($W+$jets, $t\bar{t}$)
  - $Z/\gamma^* \to \tau\tau \to \tau_h\mu$
- Common elements
  - Fake-rate technique
  - Tau embedding technique

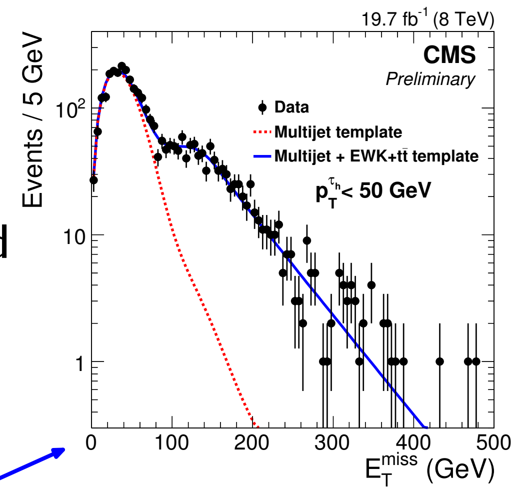
Reminder: $H^+ \to \tau\nu_\tau$, τ_h+jets final state analysis

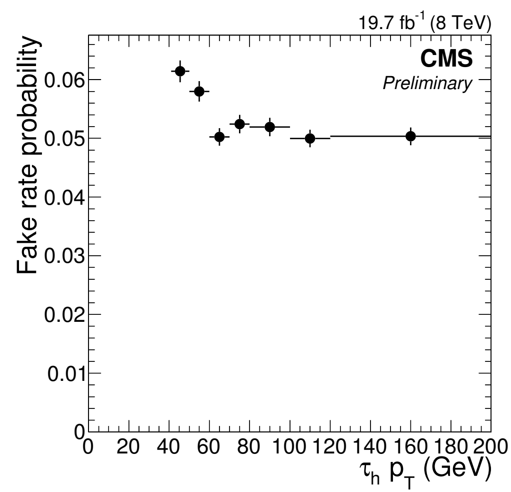


1. τ + E_T^{miss} trigger
2. τ_h identification, $p_T > 41$ GeV/c
   - Tau polarization $R_\tau > 0.7$
3. Isolated e/μ veto, $p_T > 15$ GeV/c
4. ≥ 3 hadronic jets $p_T > 30$ GeV/c
5. ≥ 1 jet b-tagged
6. Missing $E_T > 60$ GeV
7. Angular cuts between τ_h/jets and E_T^{miss}
8. Shape analysis with transverse mass $m_T(\tau_h, E_T^{miss})$ distribution
   - $m_T = \sqrt{2 p_T^{\tau_h} E_T^{miss} (1 - \cos\Delta\phi(\vec{p}_T^{\tau_h}, \vec{E}_T^{miss}))}$

Multijet background measurement

- Background from multijet events where a jet is misidentified as the τ_h
- Determine fake-rate probabilities (anti-isolated → isolated) in multijet-dominated sample
  - Before E_T^{miss}, b tagging, and τ_h-E_T^{miss} back-to-back selections
  - Performed in bins of τ_h p_T
- Non-negligible contribution from EWK+$t\bar{t}$ ⇒ fit E_T^{miss} templates
  - Multijet template from anti-isolated sample
    - ⋆ Impurity from EWK+$t\bar{t}$ (sim.) subtracted
  - EWK+$t\bar{t}$ template from simulation (isolated sample)

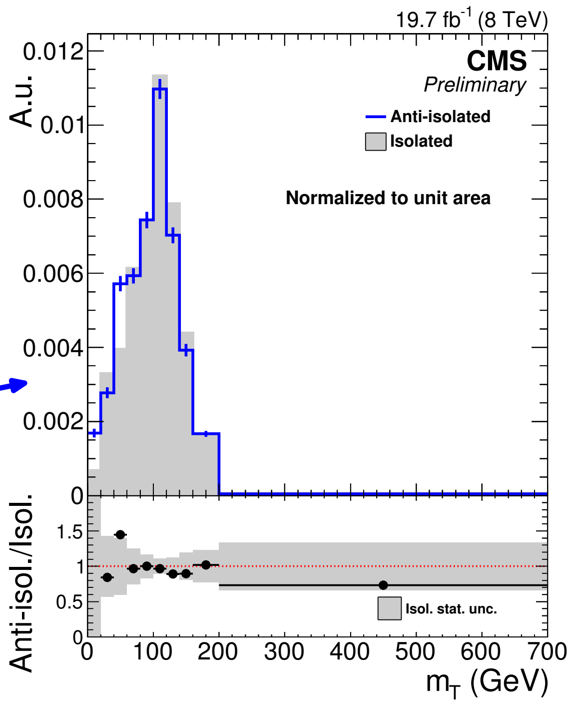




Multijet background measurement

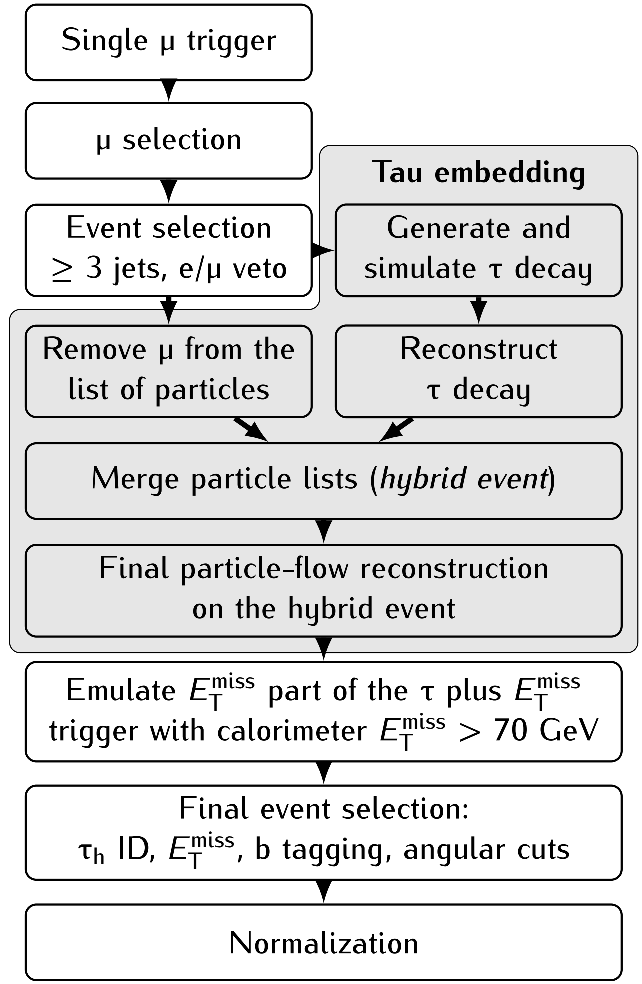
Estimate and validation

- Transverse mass distributions from anti-isolated sample after all selections in bins of τ_h p_T
  - Impurity from EWK+$t\bar{t}$ (sim.) subtracted
- Distributions weighted by the fake-rate probabilities and summed
- Dominant uncertainties
  - Simulation uncertainties affecting EWK+$t\bar{t}$ impurity (10%)
  - m_T shape difference between anti-isolated and isolated samples (5%)
    - ⋆ Evaluated with same samples as the fake rate
- Cross check by comparing the the shape of m_T distribution from anti-isolated measurement to isolated τ_h
  - Control sample by reversing b-tagging



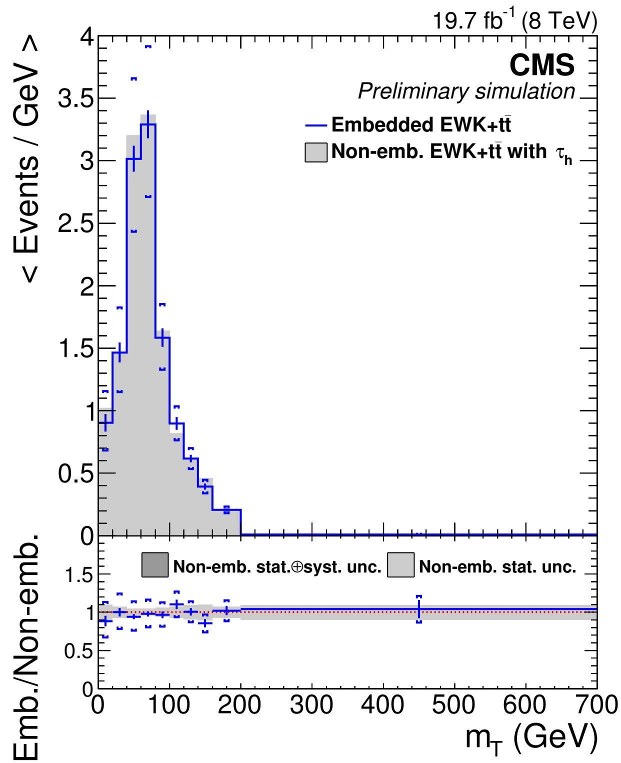
EWK+$t\bar{t}$ genuine τ background

- Background mainly from SM $t\bar{t}$, but also W + jets, Z/γ^*, single top, VV events with a genuine τ_h
- Basic idea is to exploit lepton universality $\mathcal{B}(W \to \mu) = \mathcal{B}(W \to \tau)$
- Control sample: $\mu + \geq 3$ jets
- Tau embedding done at the particle-flow level
  - Tau decay simulated and reconstructed, with tau lepton having same momentum as muon
  - Tau polarization assuming $W \to \tau\nu$ decay
- Apply remaining event selections

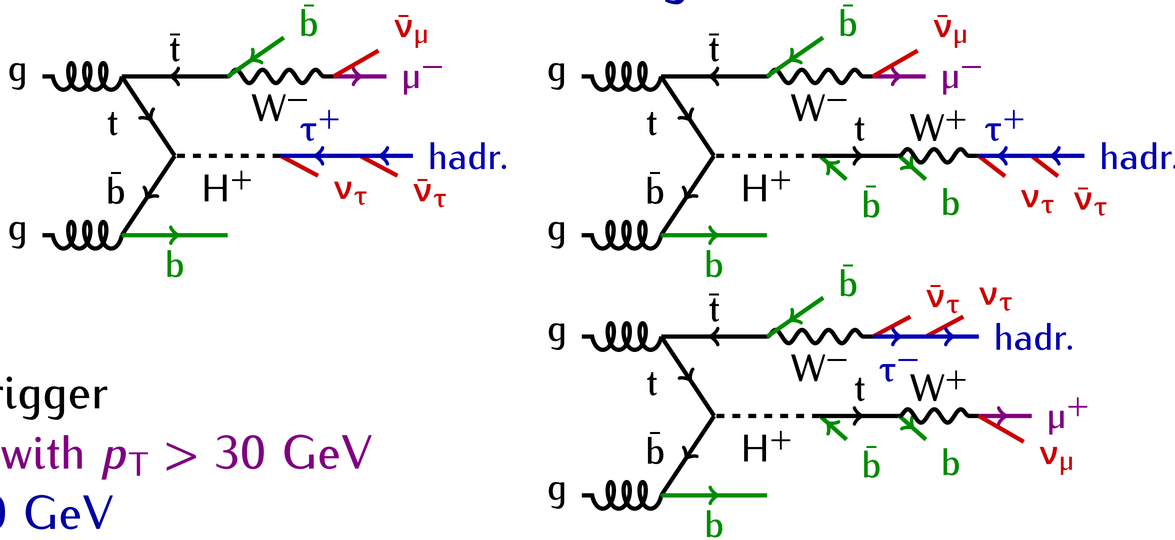


EWK+$t\bar{t}$ genuine τ background normalization and validation

- Normalization
  - $\tau + E_T^{miss}$ trigger efficiency
  - Muon trigger and ID efficiency
  - Correct for W → τ → μ events
  - Residual correction from comparing emb. and non-emb. $t\bar{t}$ simulation
- Increase statistical precision by
  - Limit only to hadronic τ decays
  - Require $p_T^{vis.\ \tau} > 30$ GeV, weight appropriately
- Dominant uncertainties
  - Residual correction (14%)
  - E_T^{miss} trigger treatment (12%)
- Validated by comparing embedded and non-embedded simulation



Reminder: $H^+ \to tb$ / $H^+ \to \tau\nu_\tau$, $\mu\tau_h$ final state analysis

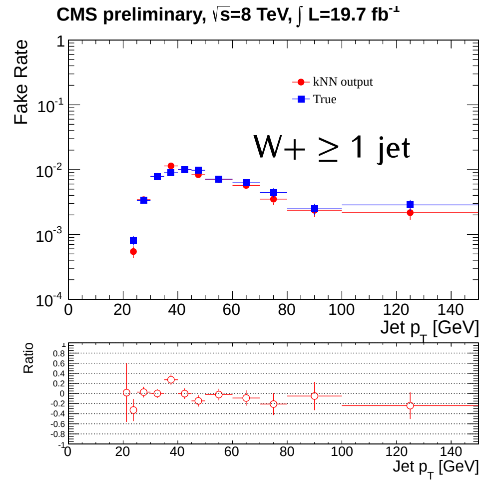


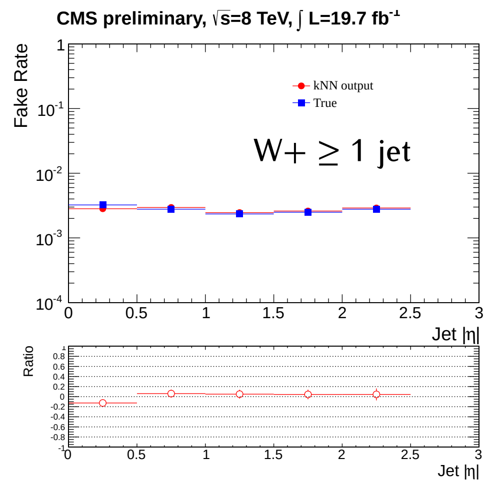
1. Single μ trigger
2. Isolated μ with $p_T > 30$ GeV
3. τ_h $p_T > 20$ GeV
4. Reject events with additional leptons
5. ≥ 2 hadronic jets $p_T > 30$ GeV
6. ≥ 1 jet b-tagged
7. Missing $E_T > 40$ GeV
8. Opposite electric charge between μ and τ_h
9. Shape analysis with b-tagged jet multiplicity

Misidentified τ_h background measurement

Fake rate

- Background from jets misidentified as τ_h
- First measure "jet $\rightarrow$ τ probability"
- From W+ $\geq$ 1 jet events
  - One isolated μ + $\geq$ 1 jet
- From multijet events
  - Single-jet triggered events with $\geq$ 2 jets
- "Jet $\rightarrow$ τ probability" parameterized as a function of the jet p_T, η, and radius ($R = \sqrt{\sigma_{\eta\eta}^2 + \sigma_{\phi\phi}^2}$)
  - Using k-Nearest Neighbour (kNN) regression





Misidentified τ_h background measurement

Estimate and validation

- Select "$\ell + \geq 3$ jet" events
  - 1 isolated e/μ + E_T^{miss} + ≥ 3 jets + ≥ 1 b-tagged jets
  - Dominated by W + jets and $t\bar{t} \rightarrow \mu$ + jets events
- Apply to every jet the "jet $\rightarrow \tau$ probability"
- Jet quark/gluon composition differ between W+jet, multijet, and "$\ell + \geq 3$ jet" samples
  - Taken into account using weights from simulation

- Validated by applying the data-driven method to simulation and comparing with expectation from simulation using generator information

Sample	MC expectation	Estimated from MC	Estimated from data
Multijet	2341 ± 61	1983	1994
W + jets		2642	2499
Unweighted average		2312(±14 %)	2246(±11 %)
Weighted average		2095(±14 %)	2145(±11 %)

Closure test within uncertainty

$Z/\gamma^* \to \tau\tau \to \tau_h\mu$ background measurement

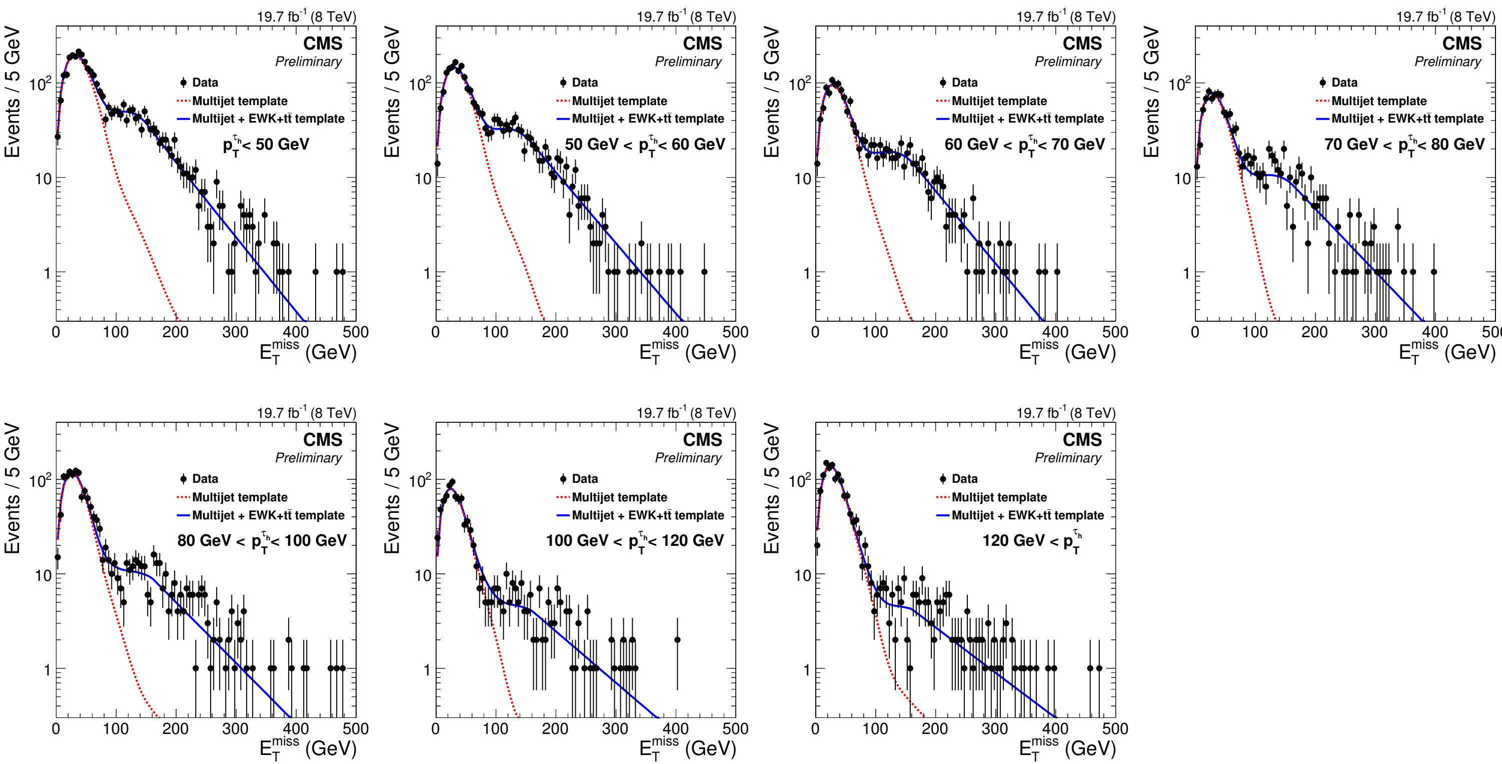
- Background from $Z/\gamma^* \to \tau\tau$ events, where
  - τ decays hadronically, identified as τ_h
  - other τ decays to a muon
- Shape of b-tagged jet multiplicity distribution from $Z/\gamma^* \to \mu\mu$ events using the tau embedding technique
  - Both muons replaced with particles from simulated tau-lepton decays
  - As in the CMS $H \to \tau\tau$ analysis
- Normalization taken from simulation
- Uncertainty on the shape from difference between embedded and simulated $Z/\gamma^* \to \tau\tau$

Summary

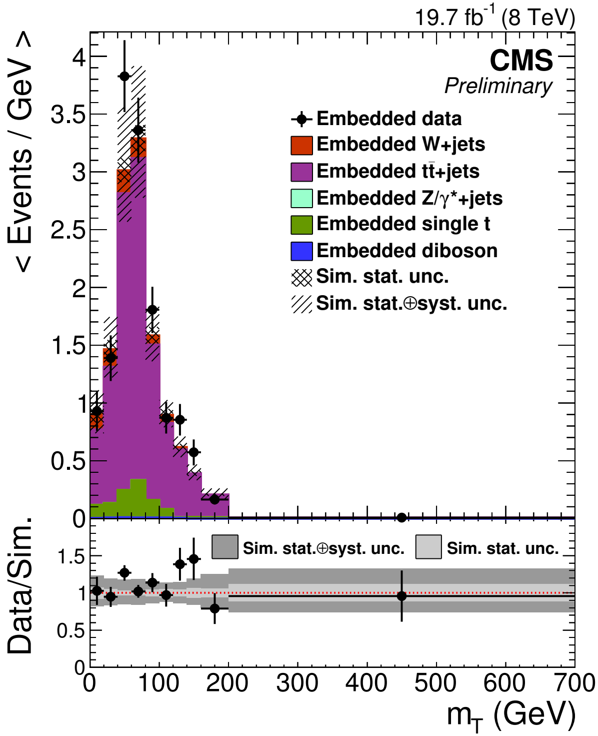
- Fake-rate technique
  - Multijet background (τ_h+jets)
    - ⋆ Fake-rate probabilities from E_T^{miss} template fits
    - ⋆ Multijet events obtained by
      - ▷ Inverting τ_h isolation
      - ▷ Weighted with the fake-rate probabilities
  - Jet misidentified as τ_h background ($\mu+\tau_h$)
    - ⋆ $\mu + \geq 3$ jets events and jet$\rightarrow\tau_h$ misidentification rate
- Tau embedding technique
  - EWK+$t\bar{t}$ genuine τ_h background (τ_h+jets)
    - ⋆ $\mu + \geq 3$ jets events and tau embedding method
    - ⋆ Normalization: correct for various efficiencies
  - $Z/\gamma^* \rightarrow \tau\tau \rightarrow \tau_h\mu$ ($\mu+\tau_h$)
    - ⋆ Shape from $Z/\gamma^* \rightarrow \mu\mu$ events and tau embedding
    - ⋆ Normalization from simulation

BACKUP MATERIAL

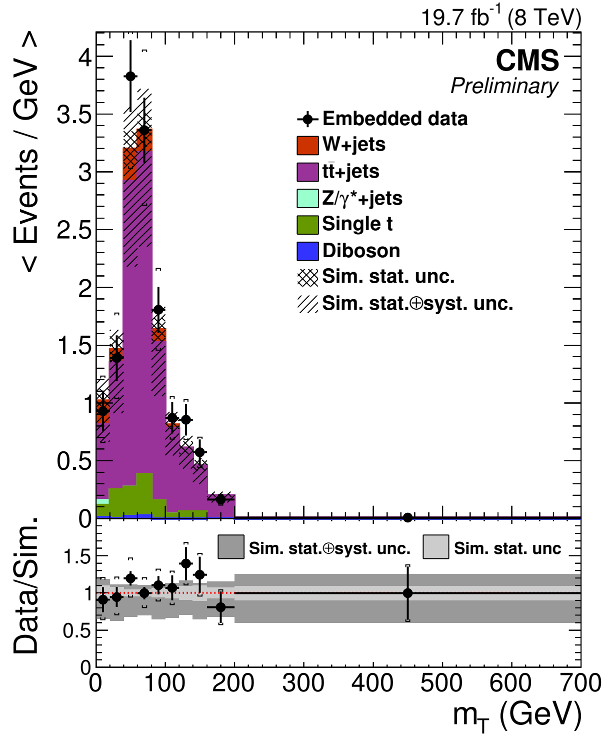
τ_h+jets: Multijet measurement normalization fits



τ_h+jets: EWK+$t\bar{t}$ genuine τ_h embedded data vs. simulation



Embedded data vs. embedded simulation



Embedded data vs. non-embedded simulation