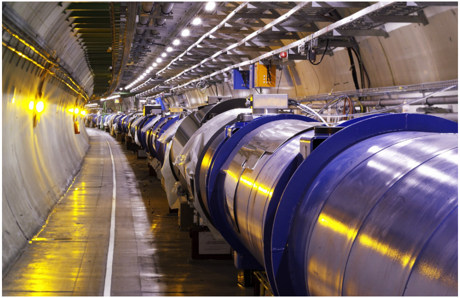
The observation of the Higgs boson in the di-tau decay mode at ATLAS

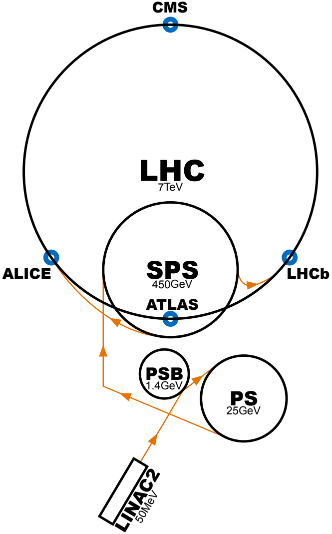
Michel Trottier-McDonald

Simon Fraser University

December 9, 2013

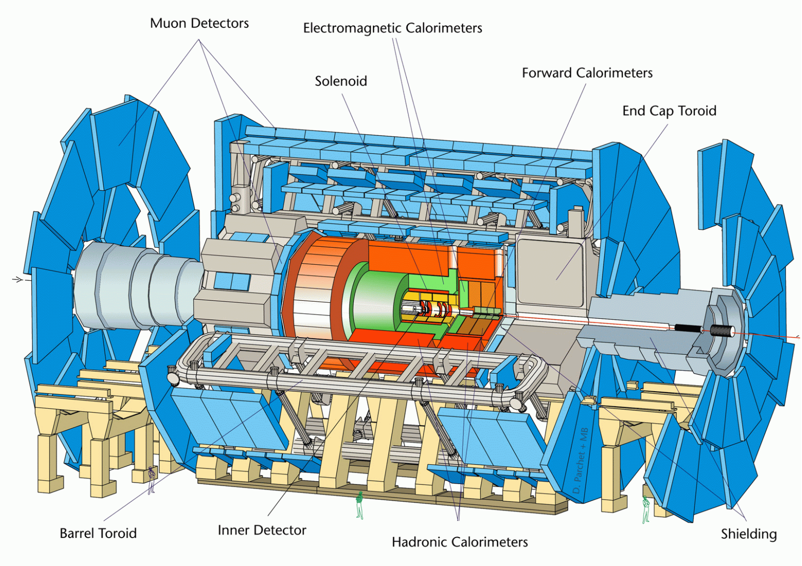
The Large Hadron Collider





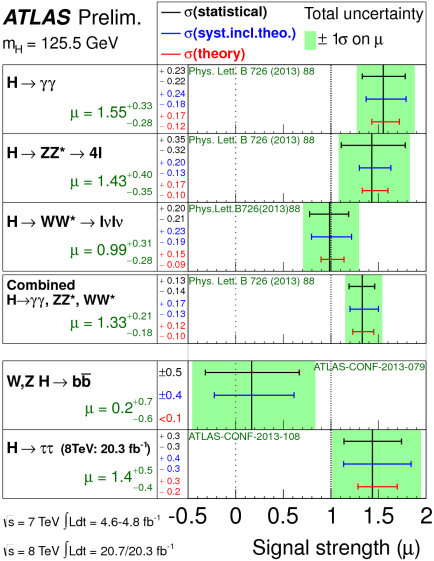
- Proton-proton collisions
- Was running at 8 TeV in **2012**:
  - 23.3 fb^{-1} (2.4×10^{15} collisions) for ATLAS
  - Bunch spacing of 50 ns
  - Peak luminosity $\sim 8 \times 10^{33}\text{cm}^2\text{s}^{-1}$

The ATLAS Detector



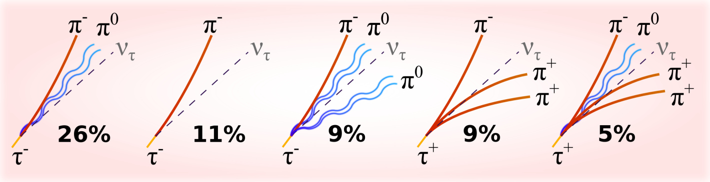
ATLAS Search for the Higgs Boson

- We have entered the measurement era for the Higgs boson!
- Among other things, it means looking for Higgs decays we haven't yet observed
- On November 26, ATLAS announced first 4.1σ evidence for $H \to \tau\tau$
- Taus have a reputation for being difficult: how did we achieve this?

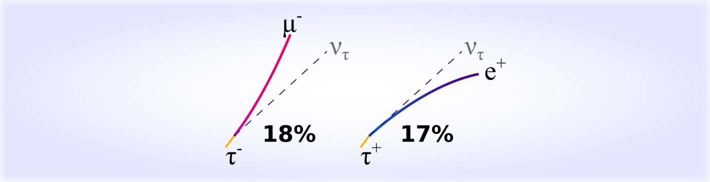


Main Tau Lepton Decay Modes

Hadronic decays

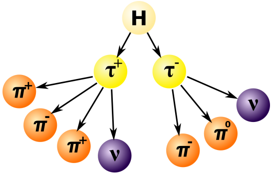


Leptonic decays

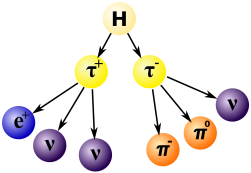


mass : 1.78 GeV **proper length : 87μm** $\pi^{\pm} \rightarrow$ **prongs**

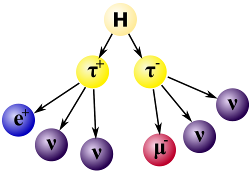
The possible final-states for $H \to \tau\tau$ are:



$\tau_{\text{had}}\tau_{\text{had}}$
hadronic-hadronic
42%



$\tau_{\text{lep}}\tau_{\text{had}}$
leptonic-hadronic
46%



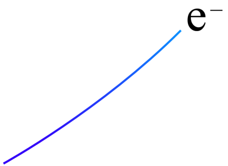
$\tau_{\text{lep}}\tau_{\text{lep}}$
leptonic-leptonic
12%

We can search for $H \to \tau\tau$ with $\tau_{\text{lep}}\tau_{\text{lep}}$ alone

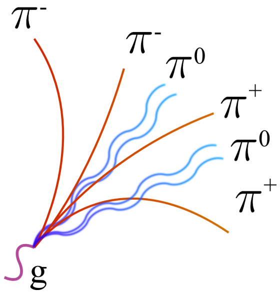
Learning to reconstruct hadronic tau decays can (in principle) buy us **8 times** more signal

Tau Lepton Backgrounds

- **Electrons**
  - electrons can look like single charged pions
  - Denser energy depositions
  - More transition radiation

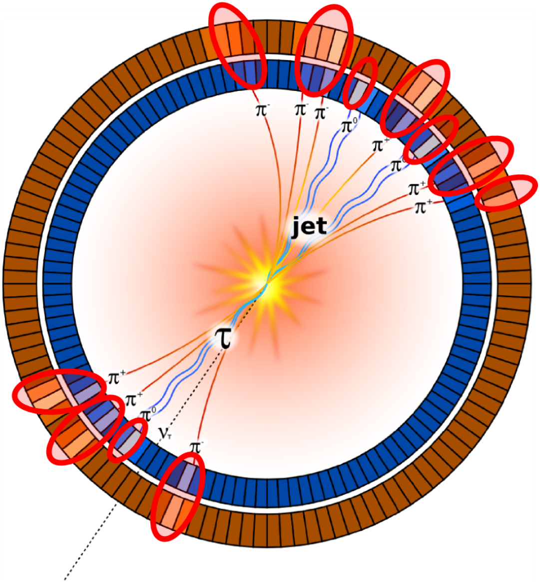


- **QCD jets**
  - Collimated showers of hadrons, can fake any hadronic tau decay
  - Most frequent type of object (by far) appearing in the detector
  - Comes directly from primary interaction vertex
  - Wider energy depositions
  - Larger number of final state particles
  - Invariant mass not resonant



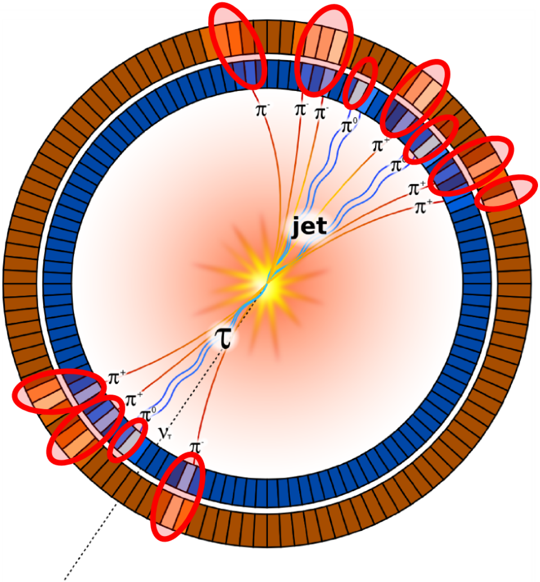
Tau Reconstruction

- The anti$-k_T$ algorithm gathers calorimeter energy deposits into objects
- We take all anti$-k_T$ outputs that fall within the tracking system range ($|\eta| > 2.5$, $p_T > 10$ GeV), and open them up to look at their substructure
- We scrutinize the configuration of the associated tracks and the energy configuration in the calorimeter



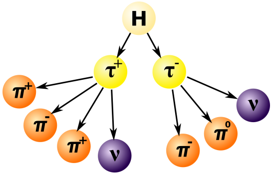
Tau Identification

- There are many observables which we reconstruct from tracking and calorimeter information, with many degrees of correlation between them
- Taking advantage of it all is currently best achieved with BDTs
- 2012 saw the beginning of π^0 cluster counting and reconstruction within taus
- This improved tau ID by 50-100%, which was crucial for $H \to \tau\tau$
- We have yet to explore the full potential of the new π^0 reconstruction

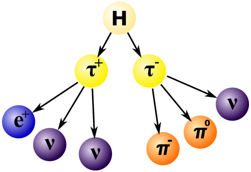


The Final State Particles (Signal)

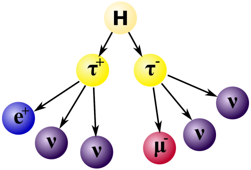
The search is divided in three **channels**:



$\tau_{\text{had}}\tau_{\text{had}}$
hadronic-hadronic



$\tau_{\text{lep}}\tau_{\text{had}}$
leptonic-hadronic



$\tau_{\text{lep}}\tau_{\text{lep}}$
leptonic-leptonic

In which we look for:

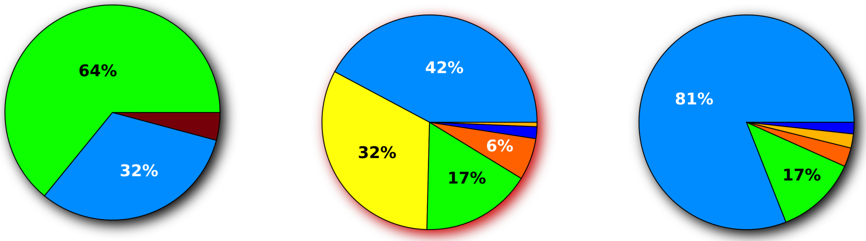
- 2 hadronic tau decays
- Any number of jets
- Missing E_T

- 1 hadronic tau decay
- 1 electron or muon
- Any number of jets
- Missing E_T

- 2 electrons or muons
- Any number of jets
- More missing E_T

We will focus on $\tau_{\text{lep}}\tau_{\text{had}}$

The Final State Particles (Backgrounds)



$\tau_{\text{had}}\tau_{\text{had}}$ | $\tau_{\text{lep}}\tau_{\text{had}}$ | $\tau_{\text{lep}}\tau_{\text{lep}}$

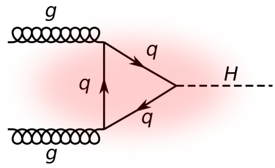
- $Z\to\tau\tau$
- QCD multijets
- $W\to\ell\nu$
- $Z\to\ell\ell$
- VV
- Top
- Others (than $Z\to\tau\tau$ and QCD)

- These object selections result in very different background compositions
- The signal is still completely overwhelmed
- We need to scrutinize event topologies for additional background rejection

Higgs Production Mechanisms

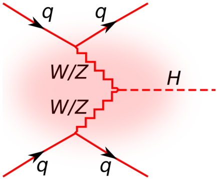
out of 2×10^{15} 8 TeV proton-proton collisions...

gluon-gluon fusion



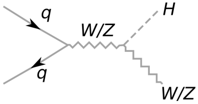
19.27 pb
(24600 $H \to \tau\tau$ events)

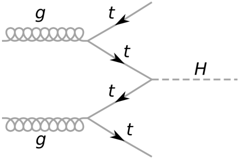
vector boson fusion



1.578 pb
(2025 $H \to \tau\tau$ events)

Higgs-strahlung
1.120 pb
(1437 $H \to \tau\tau$ events)

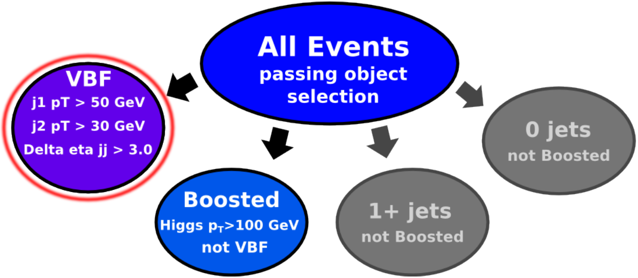




$t\bar{t}$ + Higgs
0.130 pb
(166 $H \to \tau\tau$ events)

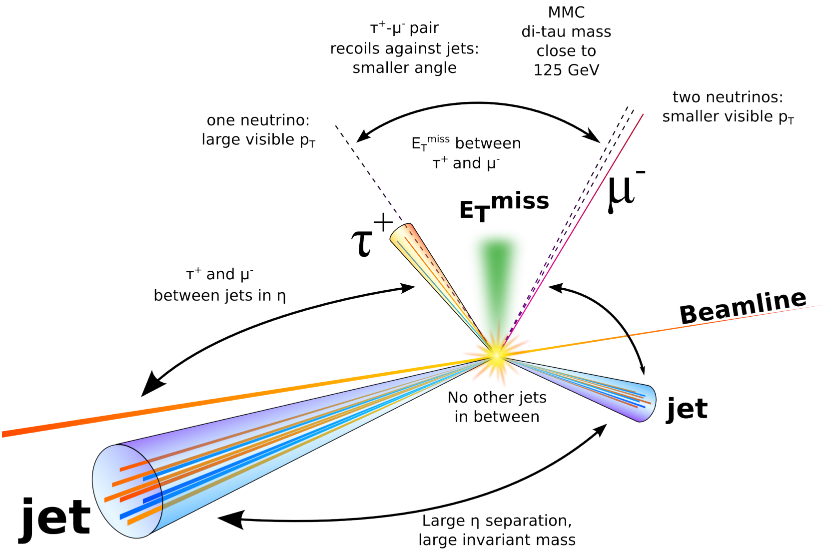
Analysis Categorization ($H \to \tau_{\text{lep}}\tau_{\text{had}}$)

- After object selection comes categorization:



- Separate the clearly distinct signal topologies to:
  - Isolate production mechanisms
  - Use variables that are relevant to each mechanism
- Few additional cuts to reject badly modelled regions and create orthogonal control regions

Topology of an Ideal VBF $H \rightarrow \tau_{\text{lep}}\tau_{had}$ Event



Quantifying the Topology (1)

- The challenge is to quantify these various features in a way that is suited to the analysis technique
- Requirements can be pretty simple (cut-based analysis):
  - $|\eta_{j1} - \eta_{j2}| > 6.0$
  - $\phi_\tau < \phi_{E_T^{miss}} < \phi_\ell$
  - $\eta_{j1} < \eta_\ell < \eta_{j2}$ or $\eta_{j1} > \eta_\ell > \eta_{j2}$
  - etc.
- **Problem is:** very few signal events look this good after full reconstruction!
- Everything is smeared by detector effects, efficiencies, pure random fluctuations, etc. If it can, it will happen...
- You want to compromise a bit on these ideal requirements in order to accept more signal, while still rejecting the backgrounds

Quantifying the Topology (2)

- Making these compromises can be a really tough optimization problem in a cut-based setting
- **Especially** on a signature with so many topological handles
- The first thing you want to do is quantify these topological handles in a continuous way, just so you have more room to optimize
- Quantifying the relative angular position of objects w.r.t. other objects required two new variables to be invented

$\phi_\tau < \phi_{E_T^{miss}} < \phi_\ell$

becomes

E_T^{miss} ϕ centrality

$\eta_{j1} < \eta_\ell < \eta_{j2}$ or $\eta_{j1} > \eta_\ell > \eta_{j2}$

becomes

ℓ η centrality

The Variables

VBF variables

$m_{\text{MMC}}(\tau_h, \tau_\ell)$

m_{j1j2}

$\eta_{j1} \times \eta_{j2}$

ℓ η centrality

$\Delta\eta_{j1j2}$

$\Delta R_{\tau\ell}$

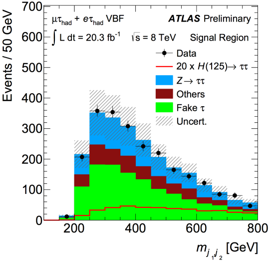
$m_T(\text{MET}, \ell)$

MET ϕ centrality

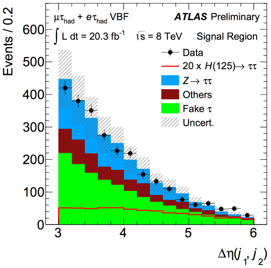
vector $\sum p_T$ (τ, ℓ, 2 lead jets)

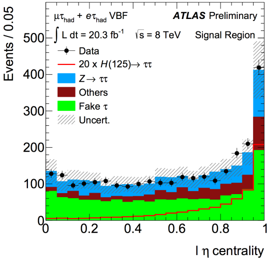
- The invariant mass of the ditau system: $m_{\text{MMC}}(\tau_h, \tau_\ell)$
- Whether the ditau system is resonant or not: $\Delta R_{\tau\ell}$
- Whether the MET and the ℓ are resonant: $m_T(\text{MET}, \ell)$
- Whether the MET originates from a boosted ditau system: MET ϕ centrality
- Presence of a third jet: vector $\sum p_T$ (τ, ℓ, 2 lead jets)
- Several features of the dijet system:
  - jets are in different hemispheres: $\eta_{j1} \times \eta_{j2}$
  - jets are well-separated in η: $\Delta\eta_{j1j2}$
  - jets are well-separated and energetic: m_{j1j2}
  - The ditau system is between the jets: ℓ η centrality

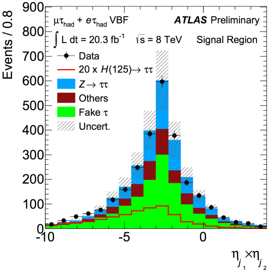
Di-jet Variables

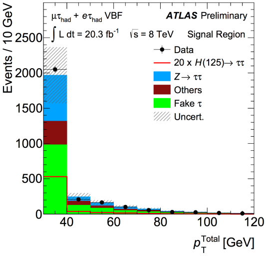


(d)

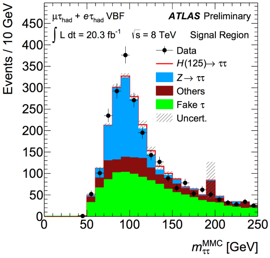


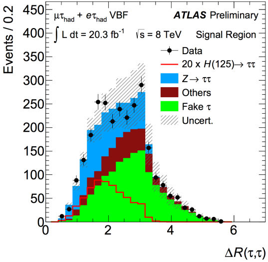


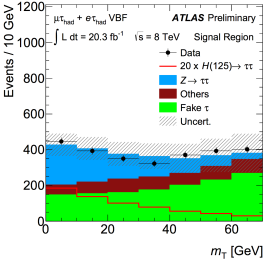


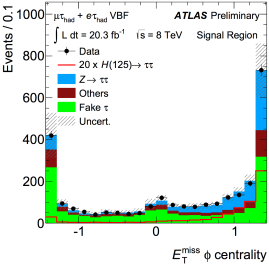


Resonance Variables





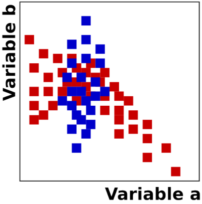


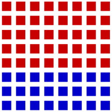


How to Best Use These Variables?

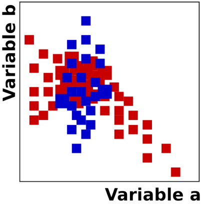
- We have a very complex variable-space with potentially interesting correlations between the variables
- These correlations may be different for signal and background: a lot of separation power may be hidden there
- This is a perfect set of conditions for machine learning algorithms such as Boosted Decision Trees (BDTs) to improve a lot over a traditional cut-based approach
- Let's take a quick look at how BDTs work

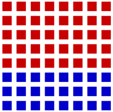
First decision tree





One boosting iteration

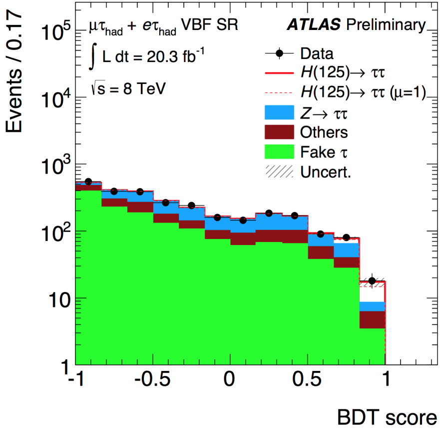




And so on...

- Every time you iterate, you create a new tree which returns its own purity score
- The final BDT score is a linear combinations of all the trees created, weighted by the classification error (more accurate trees participate more)
- The BDT creates a purity mapping of the multi-dimensional variable-space you provided
- The finer the mapping (more boosts, deeper trees), the better the classification
- The finer the mapping, the more vulnerable you are to mistake statistical fluctuations for real features in the variable-space

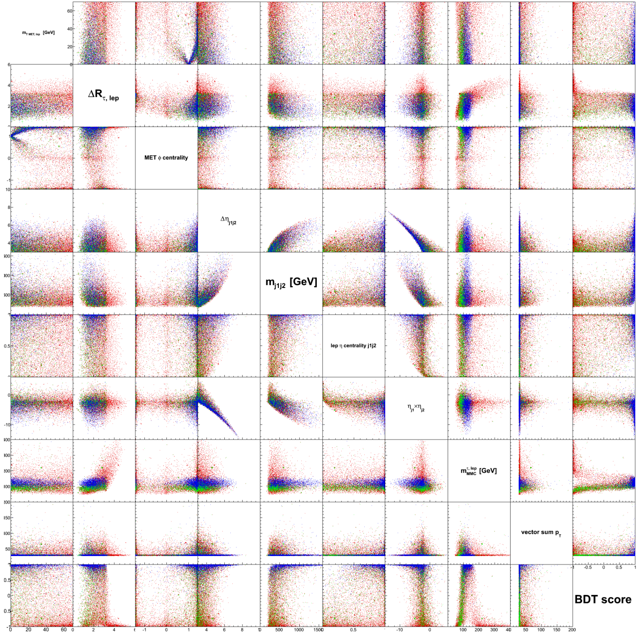
VBF Category BDT Score



Tada! A distribution in which the signal and background is more separated than in any individual input variable!

What is the BDT doing?

- Signal
- $Z \to \tau\tau$
- Other Backgrounds

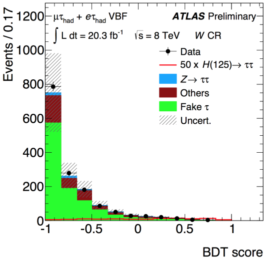


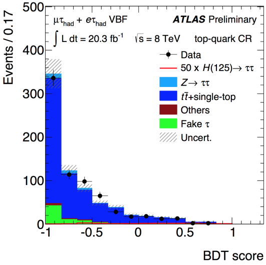
A Few Words on the Background Model

- $Z \to \tau\tau$: Hybrid data-driven/MC: Take data $Z \to \mu\mu$ events, replace muons by simulated taus, correct for acceptance and efficiencies
- $W \to \ell\mu$ and QCD : Data-driven: Take events with jets rejected by tau identification to model events where the jet passes tau identification
- $Z \to \ell\ell$: MC: Shape ($\Delta\eta$ correction) and normalization extracted from data in dedicated control regions
- Top backgrounds : MC: Normalization extracted from data in dedicated control regions
- Diboson : MC

Building Trust in the Background Model

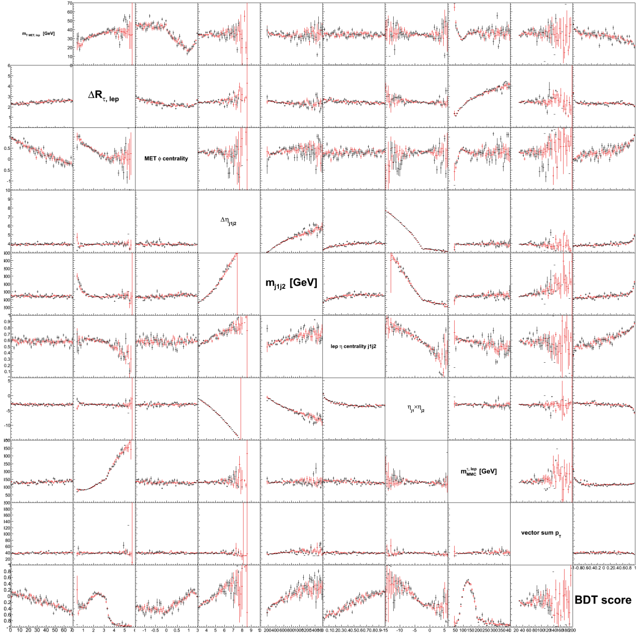
- We did a blinded analysis
- We checked modelling of all input variables at Preselection + Signal Region + Control Regions
- We unblinded the BDT score from the left going all the way up to 30% signal
- We checked the BDT score in control regions
- We explicitly checked the modelling of first order correlations





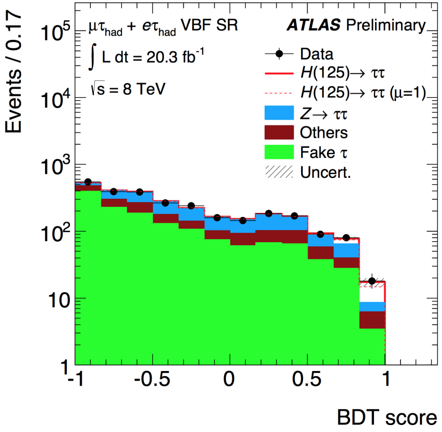
Modelling of correlations

- **Background Model**
- **Data**



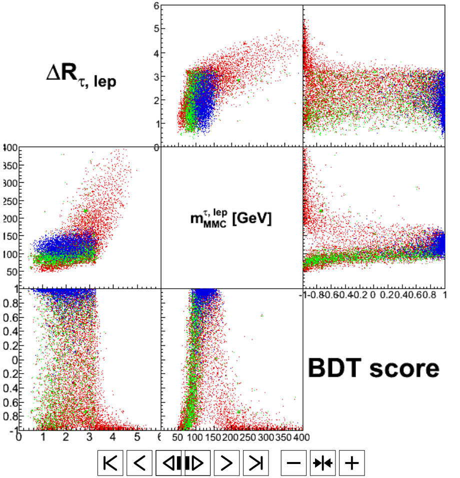
Using the BDT Score for Signal Extraction

- We fit the Background + $\mu\times$Signal model to the data using the BDT score distributions themselves
- Bins in the BDT score distribution ordered by signal purity
- Makes use of **all** signal entering the categories
- Makes use of all the variable-space directly surrounding the purest region to constrain nuisance parameters



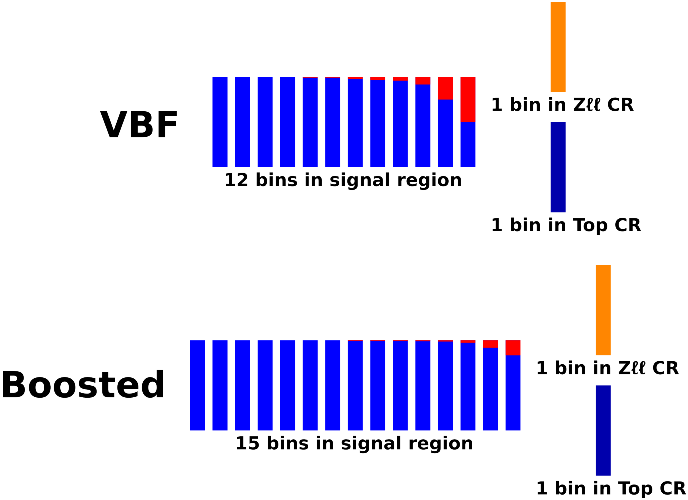
The Role of the MMC Mass

- Higgs analyses have been typically fitting the model to the data using mass distributions
- We preferred using the MMC mass as an input variable to the BDT:
  - We don't have that good a mass resolution
  - The MMC mass show correlations with other variables that differ for signal and backgrounds
- We tried a number of other ways to use the mass information outside the BDT:
  - Mass window
  - Cut on BDT score, fit mass
  - 2D BDT-mass fit



- All of which had problems, none of which outperformed our final choice

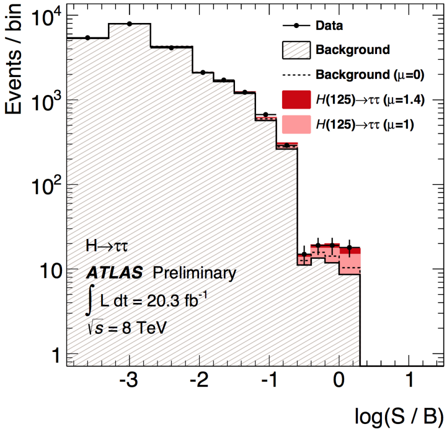
Extracting the Signal Strength



We run a simultaneous fit on all these (along with the bins from $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{lep}}$)

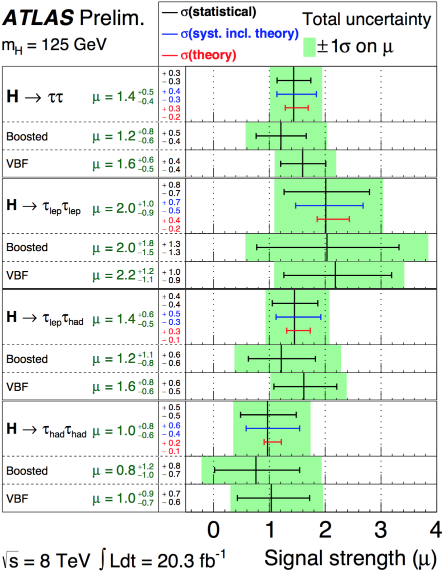
More on the Fitting Procedure

- We have shape/normalization nuisance parameters accounting for:
  - Luminosity, pileup
  - Energy scales, resolutions, efficiencies for all reconstructed objects
  - Embedding procedure
  - Various background estimation uncertainties
  - Theory cross-sections and branching ratios
- Normalization of backgrounds are left floating
- In the end, statistical uncertainties dominate the systematic uncertainties



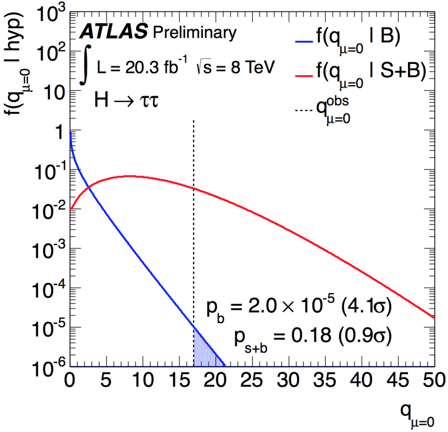
Ordering the signal region bins used in the fit by sensitivity and looking at the post-fit results

Signal Strength



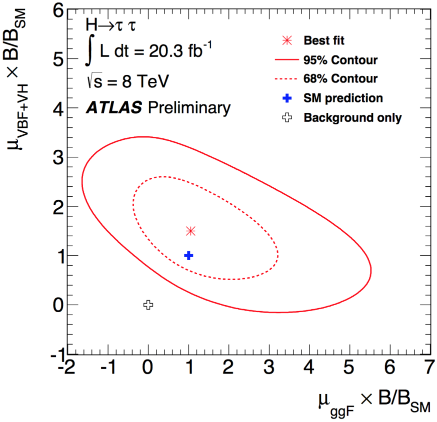
Local Significance

- Expected: 3.2σ
- Observed: 4.1σ

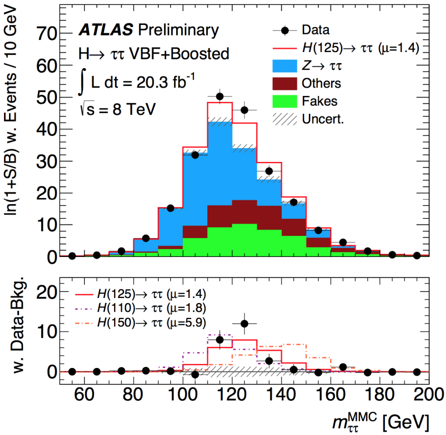


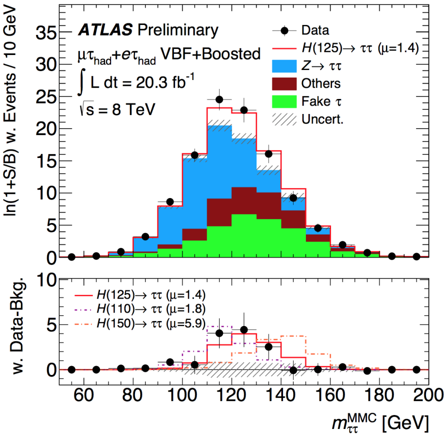
Production Cross-sections

- We are most sensitive to the VBF cross-section (good constraint to put in the ATLAS combination)
- We see a nice anti-correlation between the production modes
- Best fit sits comfortably away from null hypothesis
- Compatible with Standard Model



Weighted MMC Mass Distributions





- Weighted by $\ln(1 + S/B)$, obtained from the BDT scores
- Enhances where the 125 GeV signal is expected to be without closing down on it
- The distribution retains discrimination power between different mass hypotheses

Interpretation of the Results

- Seeing an excess in the high BDT score bins means that we see an excess of events that share most or all the features of ideal $H \rightarrow \tau\tau$ events, including the mass
- Cut-based analyses usually narrow down on a perfect signature while we also see an excess for less than perfect events, matching precisely the theory expectations (modulo the larger signal strength)
- The current analysis is a topological statement
- We now have direct evidence that the Higgs couples to leptons, and that it isn't subject to lepton universality
- There are a number of improvements on the way for the final Run I paper being discussed, the most notable being:
  - Including the 7 TeV dataset
  - Adding back the missing categories
  - (Possibly) including the VH production mode
  - Including a mass measurement
  - etc.

Conclusions

- The $H \to \tau\tau$ search at ATLAS was a problem waiting for a BDT solution
- A BDT approach required us to be much more careful about the background model, since we have to model a bigger signal region accurately
- It also got us to pay attention more closely to the treatment of systematics, which has been under intense scrutiny for several months
- Despite being preliminary, this analysis is mature and a good model to follow for model-dependent searches with complex signatures
- It paves the road for $H \to \tau\tau$ property measurements during Run II

Backup

Backup Table of Content

1. **Neutral pions in hadronic tau decays**
   - π^0 counting with BDTs
   - Reconstructing π^0 4-vectors
   - Performance
2. **Object selection**
3. **Angular centrality variables**
4. **Boosted category plots**
5. **BDTs**
6. **Techniques to take advantage of the BDT**
   - Increasing training/testing statistics
   - Taming the BDT score shape
   - Exploit all correlations
7. **Systematics**

Main steps in algorithm

First:

- Counting π^0s by looking at global tau features.

Then:

if $N_{\pi^0} > 0$:

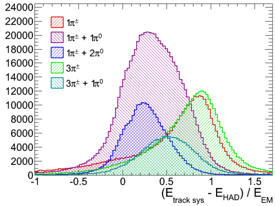
- Select the clusters in $\Delta R < 0.2$ of the track system, correct them for noise/U.E./pileup.
- Find the best cluster or pair of clusters to contain the π^0 energy.
- Correct their energy from contamination of hadronic energy if necessary.
- Sum the newly found π^0 clusters and the tracks together to reconstruct a new τ 4-vector.

else:

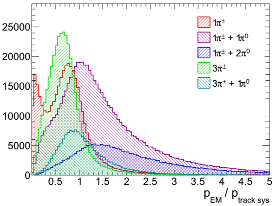
- Consider only the tracks for the new τ 4-vector.

Counting π^0s

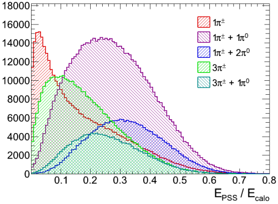
Fraction of the $\pi^{\pm}$ energy in the EM cal.



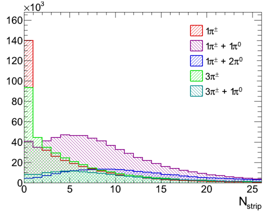
Momentum in the EM cal. over momentum of the tracks



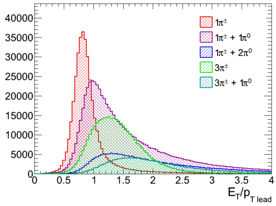
Fraction of energy in the Presampler/strip layers



Number of cells in the strip layer above 200 MeV

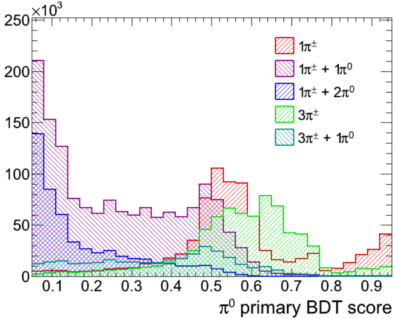


E_T over p_T of the lead track

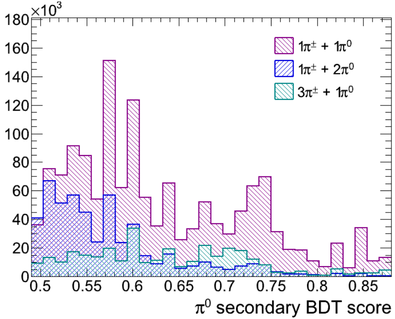


Throw them all in two BDTs

Trained on BDTmedium taus, $p_T > 20$ GeV, mc11c_7TeV, Ztautau, ZPrime, Htautauhh



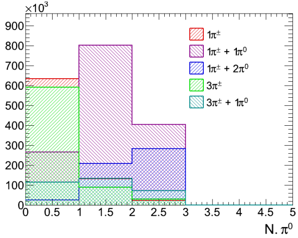
- Signal: Taus without π^0s
- Background: Taus with π^0s
- Optimal cut: 0.465
- Signal Efficiency: 81.8%
- Background Rejection: 82.4%



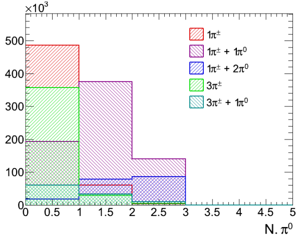
- Signal: 1p Taus with 1 π^0s
- Background: 1p Taus with 2 π^0s
- Optimal cut: 0.565
- Signal Efficiency: 54.5%
- Background Rejection: 59.8%

π^0 counting performance

- π^0/No π^0 Ratio = 1.551 (should be 1.539)
- 1 π^0/2 π^0 Ratio = 2.827 (should be 2.849)
- Distinguishing 1 from 2 π^0s is pretty hard with the existing variables.
- Potential improvements:
  - Counting conversion tracks.
  - Playing with strip cluster layers.
- The two π^0 BDTs are in the tau D3PDs: tau_calcVars_pi0BDTPrimaryScore tau_calcVars_pi0BDTSecondaryScore



All taus

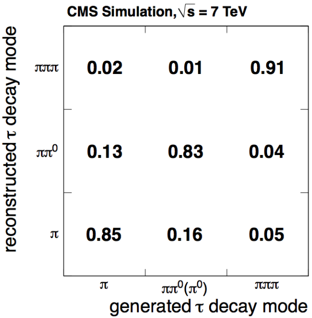


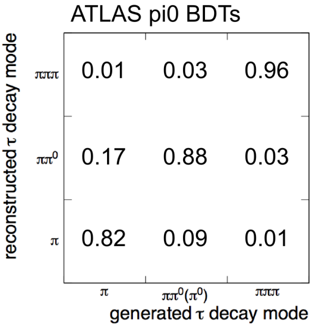
BDTmedium taus

Counting π^0s

Table: All main decays

Reco/Truth	$1\pi^\pm$	$1\pi^\pm + 1\pi^0$	$1\pi^\pm + 2\pi^0$	$3\pi^\pm$	$3\pi^\pm + 1\pi^0$
$1\pi^\pm$	81.8	12.6	4.8	0.5	0.2
$1\pi^\pm + 1\pi^0$	11.6	47.5	38.1	0.9	1.9
$1\pi^\pm + 2\pi^0$	2.6	26.5	68.9	0.5	1.4
$3\pi^\pm$	0.8	1.6	1.3	92.0	4.2
$3\pi^\pm + 1\pi^0$	0.3	1.2	3.5	50.2	44.9





Pi0Finder: Reconstruction of π^0 4-vectors

- Correction for noise/U.E/pileup:
  - Use the ring of $0.2 < \Delta R < 0.4$ around the track system to estimate the noise/U.E./pileup activity.
  - Subtract this activity from the clusters found in $\Delta R < 0.2$.
  - Keep only the clusters within $\Delta R < 0.2$ that survives the correction.
- Then pick the cluster **or** the pair of clusters with the largest:

$$\pi^0 \text{ likeness score} = \frac{E^{\text{PSS}}_{\text{cluster(s)}}}{f^{\text{HAD}}_{\text{cluster(s)}} + x\sqrt{\left|\frac{E_{\text{cluster(s)}}}{E_{\text{calo}} - E_{\text{tracks}}}\right|}}$$

- Apply a hadronic energy contamination correction if necessary.

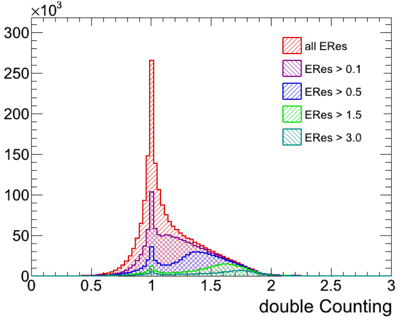
The code and documentaion (README file) are in SVN.

**I tried reducing the weight of clusters found close to tracks ($1/\Delta R^2$), but no improvement could be seen. It got worse at higher p_T^τ.*

Hadronic energy contamination correction

Finding out if there is need for correction:

- double Counting = $E_{\text{tracks} + \pi^0 \text{ clusters}} / E_{\text{calo}}$ is strongly correlated with π^0 energy resolution.
- If double Counting > 1.05, the correction gradually kick in.



The corrected energy takes the form:

$$E^{\text{cluster}}_{\text{corrected},1} = \frac{aE^{\text{cluster}}_{\text{PSS}} + bE^{\text{cluster}}_{\text{Layer2}}}{\text{double Counting}}, \qquad E^{\text{cluster}}_{\text{corrected},2} = E^{\text{cluster}}_{\text{uncorrected}} - \frac{E^{\text{cluster}}_{\text{HAD}}}{E^{\text{calo}}_{\text{HAD}} / E^{\text{trkSys}}}$$

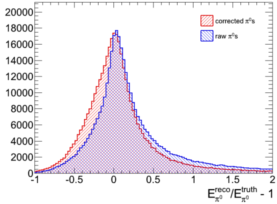
$$E^{\text{cluster}}_{\text{corrected}} = \left(1 - \frac{E^{\text{cluster}}_{\text{HAD}}}{E^{\text{calo}}_{\text{HAD}}}\right) E^{\text{cluster}}_{\text{corrected},1} + \left(\frac{E^{\text{cluster}}_{\text{HAD}}}{E^{\text{calo}}_{\text{HAD}}}\right) E^{\text{cluster}}_{\text{corrected},2}$$

**I tried also correcting the cluster direction during this correction, but performance only degrades.*

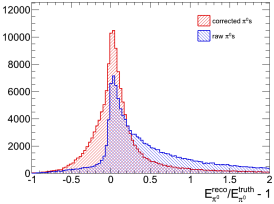
How well it works

π^0 energy resolution, before and after correction:

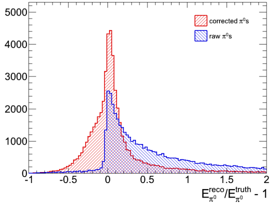
$20 < p_T^\tau < 30$ GeV



$70 < p_T^\tau < 100$ GeV



$200 < p_T^\tau < 250$ GeV



- high p_T taus are more collimated $\rightarrow$ more cluster merging $\rightarrow$ more contamination in chosen π^0 clusters.
- The contamination correction is almost unnecessary at low p_T but very important at high p_T.

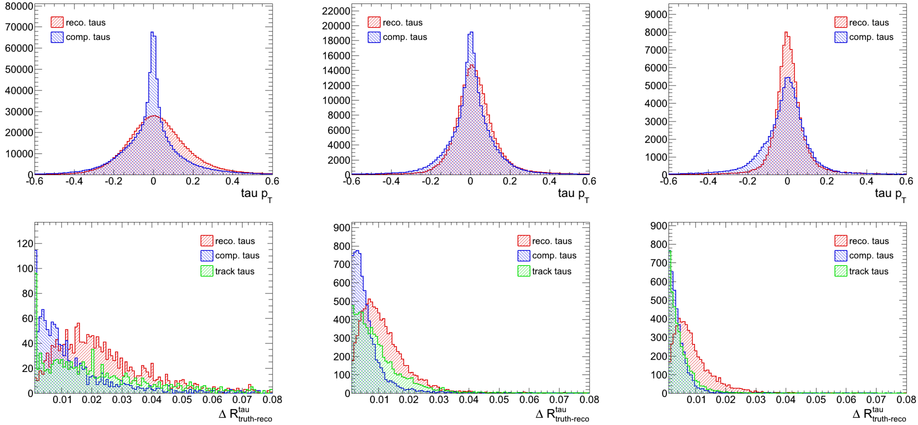
Tau 4-vector reconstruction (p_T and direction)

new taus (tracks + π^0 clusters)

tauRec taus

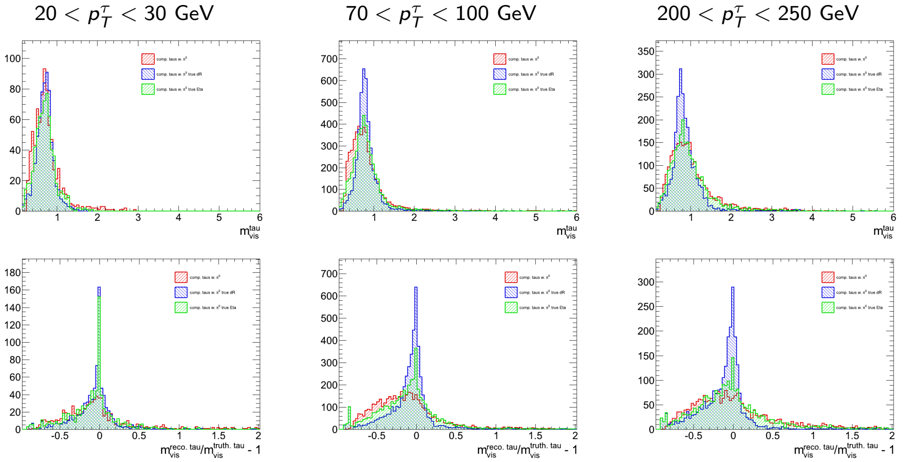
track taus (tracks only)

$20 < p_T^\tau < 30$ GeV | $70 < p_T^\tau < 100$ GeV | $200 < p_T^\tau < 250$ GeV



Tau invariant mass (mass and resolution)

completely reconstructed π^0s
reconstructed π^0s with true η
reconstructed π^0s with true η and ϕ



Comments on tau 4-vector reconstruction

- This is only when taking the cluster corrected directions w.r.t. tau vertex (atTJVA). Otherwise, artificial distance in η is created between tracks and clusters.
- The new reconstructed tau p_T resolution appears better mostly because of $\pi^{\pm}$-only tau decays where only the track is taken.
- At high p_T, the calo-only p_T still does better.
- The new tau direction is better.

Object selection (1)

Muons:
track quality
isCombinedMuon
$|\eta_\mu| < 2.5$
isolated
$p_T > 17$GeV (2011 LTT)
$p_T > 20$GeV (2011 SLT)
$p_T > 17$GeV (2012 LTT)
$p_T > 26$GeV (2012 SLT)

Jets:
$|\eta_j| < 4.5$
$p_T > 25$GeV (2011)
$p_T > 30$GeV (2012)
if $|\eta_j| < 2.4$:
jvf > 0.75 (2011)
jvf > 0.5 (2012)

Electrons:
object quality
TightPP
$|\eta_e| < 2.47$
$1.37 > |\eta_e| > 1.52$
isolated
$p_T > 17$GeV (2011 LTT)
$p_T > 25$GeV (2011 SLT)
$p_T > 20$GeV (2012 LTT)
$p_T > 26$GeV (2012 SLT)

Taus:
$N^\tau_{\text{trk}} = 1||3$
author $= 1||3$
BDTMedium
EleBDTMedium $= 0$
MuonVeto $= 0$
$|\text{charge}| = 1$
$|\eta_\tau| < 2.5$
$p_T > 25$GeV (2011 LTT)
$p_T > 20$GeV (2011 SLT)
$p_T > 25$GeV (2012 LTT)
$p_T > 20$GeV (2012 SLT)

Object selection (2)

The full object selection/preselection for 2011 is documented here.
Modifications for 2012 are documented here.

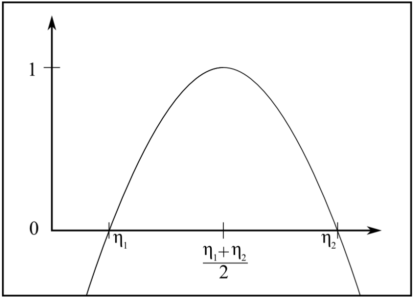
Isolation criteria is on both tracks and calorimeter energy:

- **Muon isolation:**
  - track: mu_ptcone40/mu_pt $\leq$ 0.06
  - calo: mu_etcone20/mu_pt $\leq$ 0.04 for 2011 (0.06 for 2012)
- **Electron isolation:**
  - track: el_ptcone40/el_cl_pt $\leq$ 0.06
  - calo: el_Etcone20/el_cl_pt $<$ 0.08 for 2011 (el_topoEtcone20/el_cl_pt $<$ 0.06 for 2012)

- **dilepton veto**: Considered part of the object selection. Reject any event with more than one lepton after full lepton selection.
- **Overlap removal**: In the following order: $\mu > e > \tau >$ jet with the overlap criterion being $\Delta R < 0.2$.

η centrality variables

Goal: make a continuous quantity out of $\eta_{j1} < (\eta_\tau, \eta_\mu) < \eta_{j2}$:



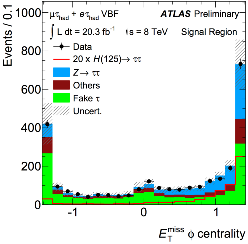
and then take the exponential of that:

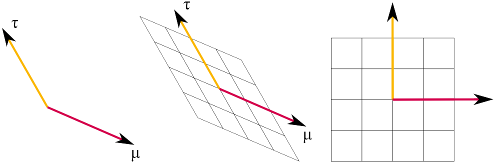
$$C_{\eta 1 \eta 2}(\eta) = \exp\left[\frac{-1}{(\eta_1 - \eta_2)^2}\left(\eta - \frac{\eta_1 + \eta_2}{2}\right)^2\right]$$

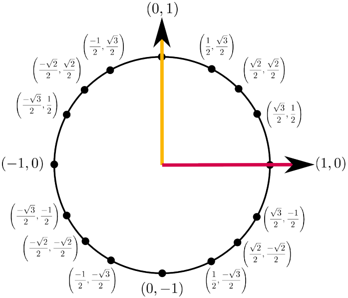
where η_1 and η_2 are the η of the VBF jets.

- $C_{\eta 1 \eta 2}(\eta) = 1$ means the object is right in-between the two VBF jets.
- $C_{\eta 1 \eta 2}(\eta) = 1/e$ means the object is aligned with one of the VBF jets.
- $C_{\eta 1 \eta 2}(\eta) < 1/e$ means the object is outside the VBF jets.

Missing E_T ϕ centrality







$C_{E_T^{miss}(\phi)}$ = sum of MET unit vector components in (τ, μ)-transformed coordinates

NEW: Missing E_T ϕ centrality

- $C_{E_T^{miss}(\phi)} = \sqrt{2}$: MET in-between tau and muon
- $C_{E_T^{miss}(\phi)} = 1$: MET aligned with tau or muon
- $C_{E_T^{miss}(\phi)} < 1$: MET outside tau and muon

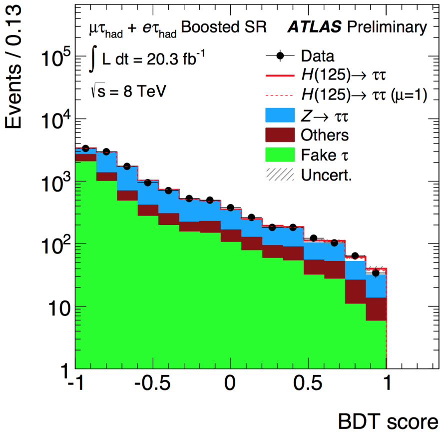
MET ϕ centrality vs. $\sum \Delta\phi$

$$\langle \phi_{\text{MET}} \rangle_{ab} \begin{cases} = \sqrt{2} & : \text{MET is exactly between } a \text{ and } b \\ > 1 & : \text{MET is somewhere between } a \text{ and } b \\ = 1 & : \text{MET is aligned with } a \text{ or } b \\ < 1 & : \text{MET is outside } a \text{ and } b \\ = -\sqrt{2} & : \text{MET is exactly opposite to the middle of } a \text{ and } b \end{cases}$$

$$\sum \Delta\phi \begin{cases} = \Delta\phi_{ab} & : \text{MET is exactly between } a \text{ and } b \\ = \Delta\phi_{ab} & : \text{MET is somewhere between } a \text{ and } b \\ = \Delta\phi_{ab} & : \text{MET is aligned with } a \text{ or } b \\ > \Delta\phi_{ab} & : \text{MET is outside } a \text{ and } b \\ > \Delta\phi_{ab} & : \text{MET is exactly opposite to the middle of } a \text{ and } b \end{cases}$$

- MET ϕ centrality uniquely specify more configurations than $\sum \Delta\phi$.
- You can divide $\sum \Delta\phi$ by $\Delta\phi$ but you will still only distinguish 2 configurations out of 5.
- A signed version of $\sum \Delta\phi$ cannot distinguish between the MET being between the two objects or being opposite to the two objects.

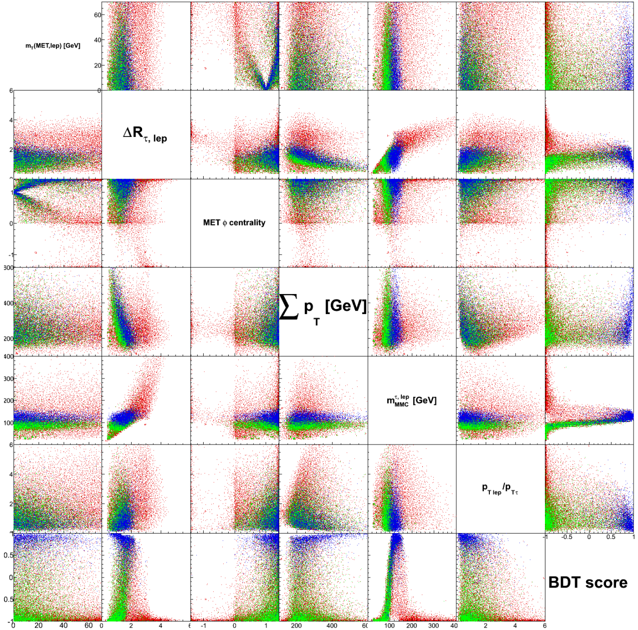
Boosted Category BDT Score



What is the (boosted) BDT doing?

- Signal
- $Z \to \tau\tau$
- Other Backgrounds

- Use the buttons at the bottom to scan through the score and see what happens when narrowing down further and further on the signal region (works in Adobe Reader).



BDT Weaknesses

BDTs are hungry beasts...

- Training sample size is a major limiting factor to the performance of a BDT: you want to have as many training events as possible
- A BDT is **always** biased to some extent on the training sample
- The best practice is to use different events for training than for the background model and signal models
- More generally, you need to know precisely what kind of signal you are looking for. A little creativity is required to use BDTs in model-independent searches

BDT Strengths

BDTs follow the physics...

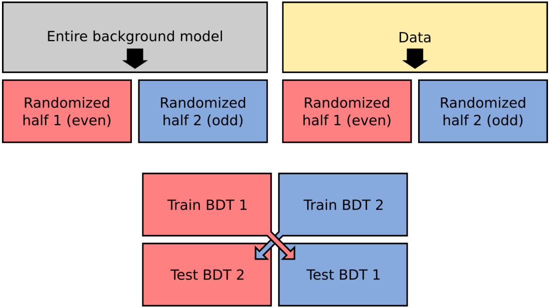
- ... provided that the variable-space handed in (the input variables) is carefully designed to quantify the features that distinguish the signal
- They are really good at finding regions of high signal purity:
  - They can unfold really complex correlations that would be very difficult to parametrize
  - You don't have to spend time optimizing cuts
  - They can teach you a thing or two about what makes the signal unique by identifying unexpected high purity regions
- You don't have to define a "signal region", as the boundary between signal and background is continuous

Training and testing

- A BDT is built by learning about the signal and background configurations. It requires a **training sample**.
- In this type of learning, bias on the training sample is unavoidable.
- The signal and background samples used for the model when evaluating limits, p-values, etc. **have to be independent** of the training sample.
- The common strategy is to split the samples in halves, use one half for training, another for testing.
- This will raise the statistical error on the model, so any trick to reduce this error again have to be deployed.

Trick 1: Cross-Evaluation

- Allows you to use the full background model when evaluating limits, p-values, etc.
- What you need:
  - samples split in a truly randomized way, there should be no bias when evaluating data (**check**).
  - to have BDT shapes similar enough for the signal to accumulate in the same place (**check**).



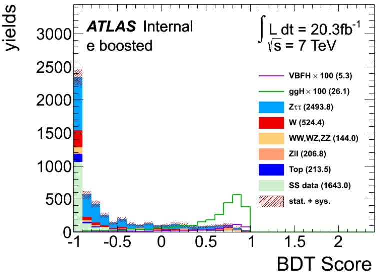
- What really matters is that both halves (especially for the data) are properly randomized. Having identical shapes is not mandatory.
- Having identical shapes is nevertheless desirable. You do not want to smear the separation.

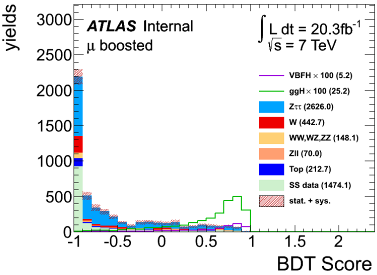
Trick 2: Different samples for training/evaluation

- For $Z \to \tau\tau$, we have both Alpgen MC and embedding samples. Embedding is used entirely for evaluation, Alpgen is used entirely for training.
- For physics processes with fake taus, we have two background estimates using independent samples:
  - Same Sign (SS) events: models QCD, and part of W+jets
  - Anti-tau ID events: the Fake Factor (FF) method, models QCD, all of W
- FF method is not available for all categories, but when it is, it offers 10x more events than SS.
- Use SS for training and FF for evaluating whenever possible.
- Use of embedding and FF haven't been shown yet: coming soon.

Trick 3: Merge electron and muon channels

- Background shapes and signal shapes are similar enough in e and μ channels that this can be done at no cost to the final sensitivity.
- This is mostly helpful when monitoring the BDT score shape in control regions.



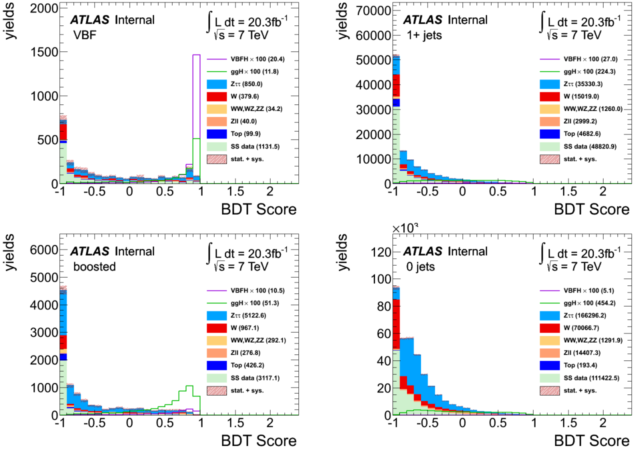


Trick 4: VBF-filtered samples

- We prepared special MC samples designed to deliver a big sample size in the VBF category.
- The events are selected from generator information, to have a VBF-like topology ($\Delta\eta_{jj}$ and m_{jj} cuts, on the leading jets)
- We made these for Z+jets and W+jets events.
- We make a patchwork of samples using these that ends up being inclusive, with the largest number of events where it is most needed.

Trick 5: Adjust the BDT score shape

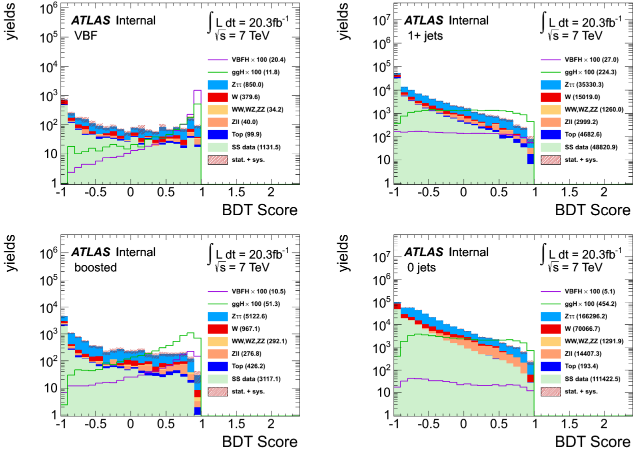
- We use *gradient boosting*, as opposed to the more traditional Adaboost.
- The BDT score shape is more manageable: signal and background always overlap, with no cost to separation power.
- Small background model sample size more manageable.
- Makes signal region easier to control.



Trick 5: Adjust the BDT score shape

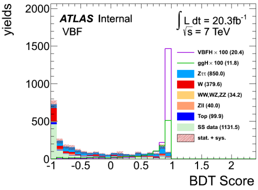
(**Same plots in log scale**)

- BDT parameters have also been optimized for separation and score shape.

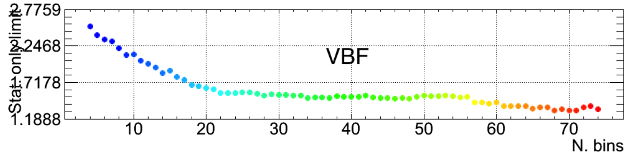


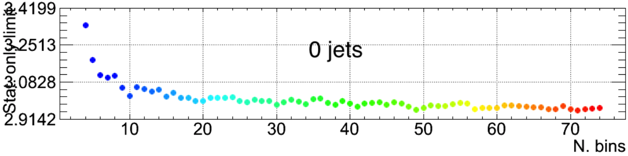
Trick 6: Binning optimization

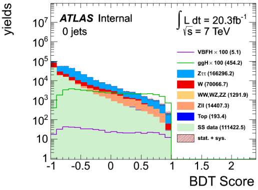
- We have been doing several binning scans, with the Kyle Fix fully activated on all backgrounds.
- The gradient boosting BDT shape makes this optimization both easy and tricky.
- You really want to take advantage of that fast rise in the VBF BDT score.
- If the binning is too fine, the limit calculation takes a lot of time.



- A non-uniform binning is best, but not easy to find.







Trick 7: MMC mass as an input variable

- This is by far the most controversial aspect of the analysis, but convenors, Ed. board members, etc. have been supportive so far (with a healthy amount of skepticism).
- **Pro :** All the signal in the category is available for the limit extraction. It is ordered by purity along the score. This is like using the full series expansion instead of just the 1^{st} order term.
- **Pro :** All the background is there too. Great for constraining systematics and mitigating small sample size problem.
- **Pro :** It enhances the separation power of the BDT significantly.
- **Con :** difficult to produce a mass spectrum with expected localized excess which is a very valuable PR plot. There is a way around though.
- **Con :** A bit more difficult to control the MMC mass distribution, but not impossible.

Trick 8: Patch remaining holes in the BDT score

- Some backgrounds were previously poorly represented in the high-BDT score region.
- HistFactory assumes there is no background in one bin if the yield is 0.
- This is a bad assumption if the equivalent luminosity of the model sample is smaller than the real data luminosity.
- Kyle Cranmer provided a statistically correct treatment (the so-called *Kyle Fix*), which consist of adding background events in these bins, with specified weight and error.
- This is conservative, as it assumes a flat background shape.
- Our signal region is now statistically strong, so that this is almost not needed anymore.

Nuisance Parameters

- We have algorithms to:
  - Prune statistically challenged shape variations
  - Smooth shape variations
- In the end, even the largest systematic error does not influence the result much, because the sensitivity comes from bins with small number of events (~ 15)