SM H→$\tau\tau$ Search in ATLAS:

2 analysis strategies

Daniele Zanzi (MPP)

ATLAS

FSP 101

ATLAS

$\Delta p \cdot \Delta q \geq \frac{1}{2}\hbar$

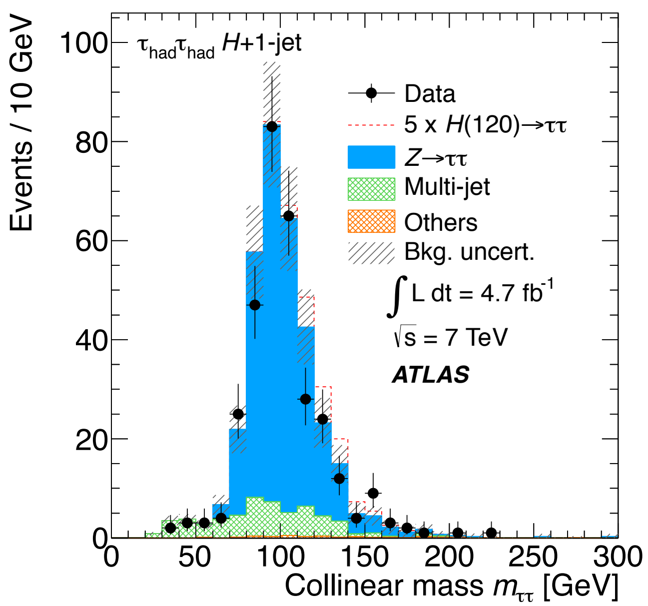
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

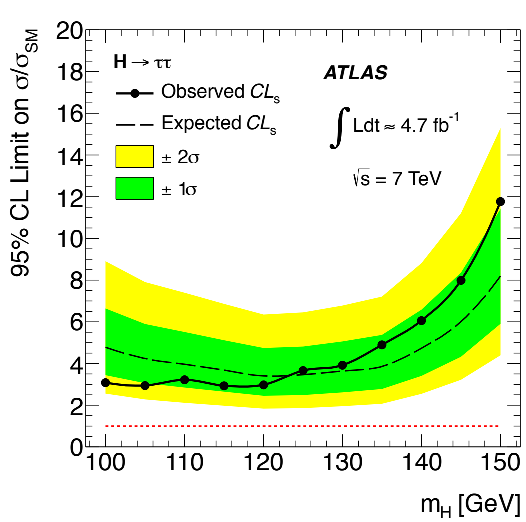
SM H→ττ Search

March 2012	October 2012	November 2013

March 2012

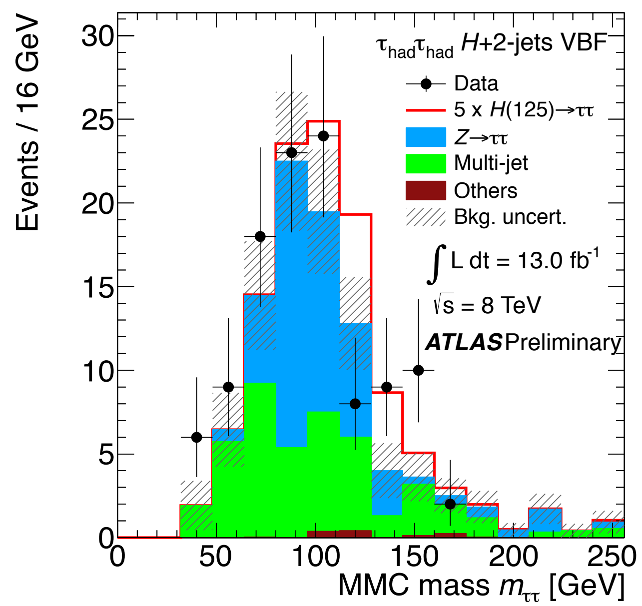
- Cut-Based analysis
- 5 fb^{-1}
- Exclusion upper limits at $2.9\times\sigma_{SM}$

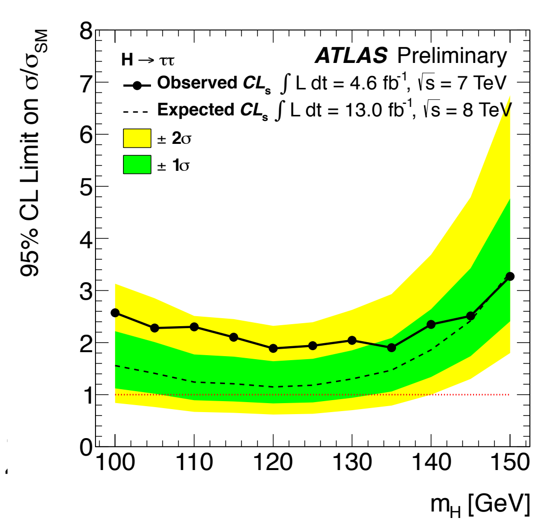




October 2012

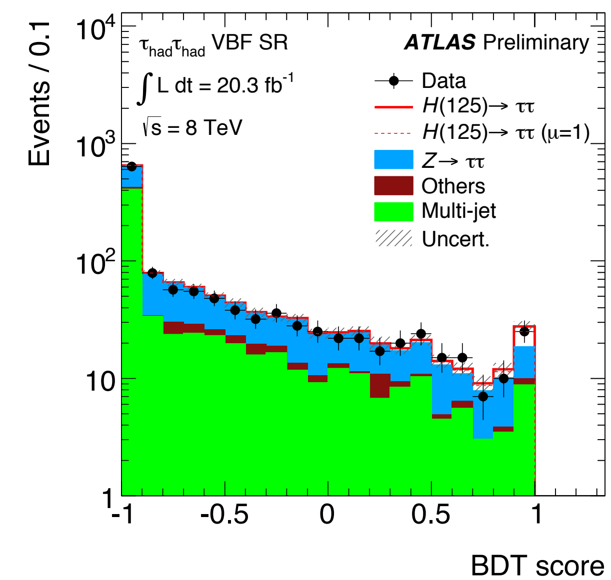
- Cut-Based analysis
- 5+13 fb^{-1}
- Exclusion upper limits at $1.9\times\sigma_{SM}$

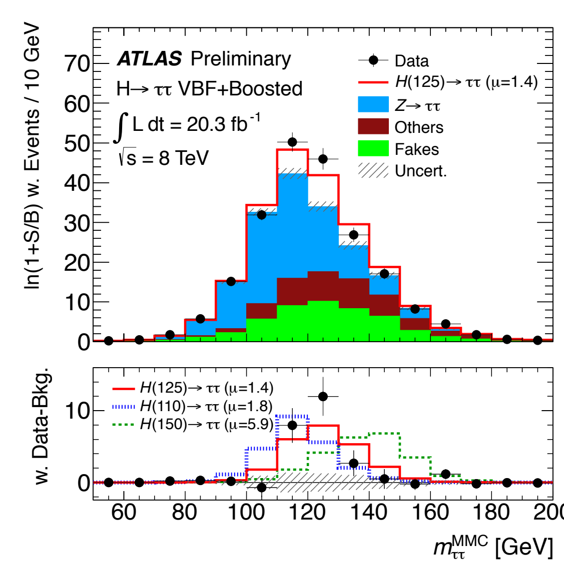




November 2013

- Multi-Variate analysis
- 20 fb^{-1}
- First evidence of Higgs decaying to fermions with 4.1σ significance





Outline

- What's a Search?
- Analysis Strategies
- Case Study: SM H$\rightarrow\tau\tau$
- Strategy Comparison:
  - Sensitivity
  - Systematic Uncertainty (Theory, Experimental)
- Conclusions

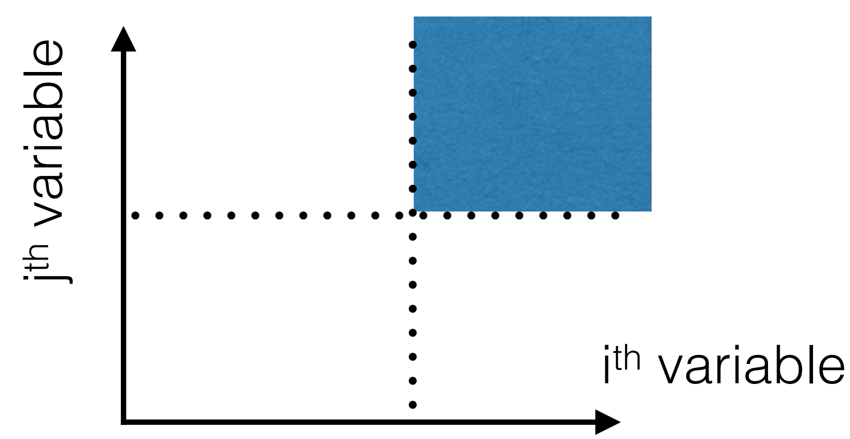
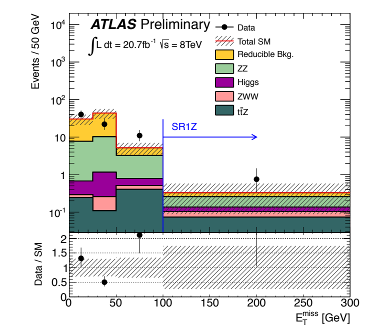
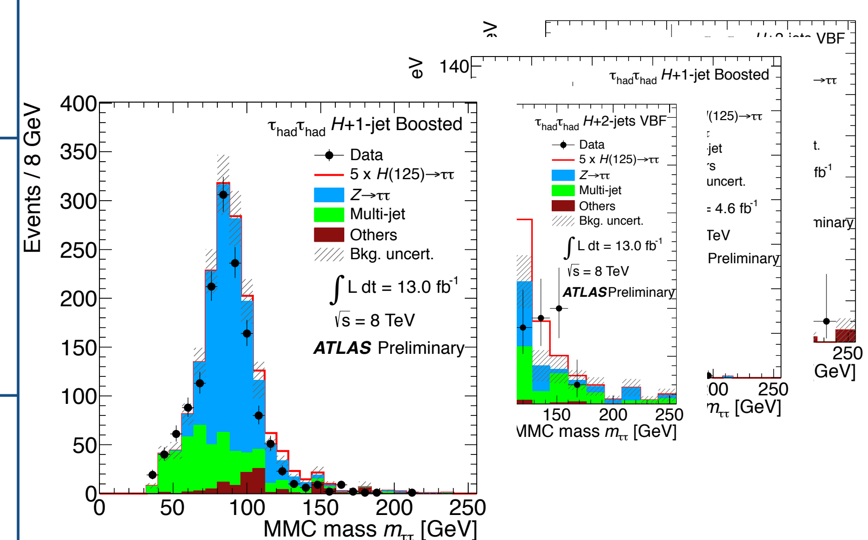
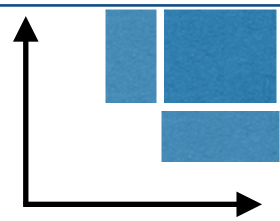
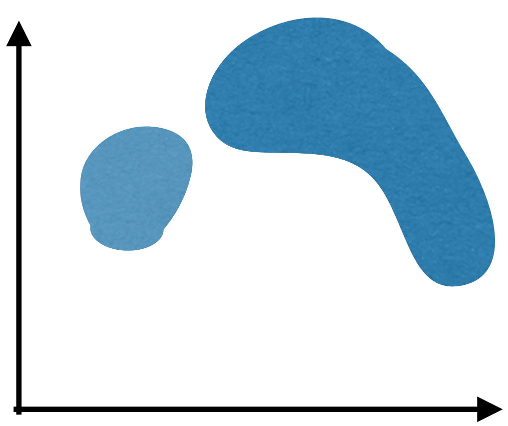
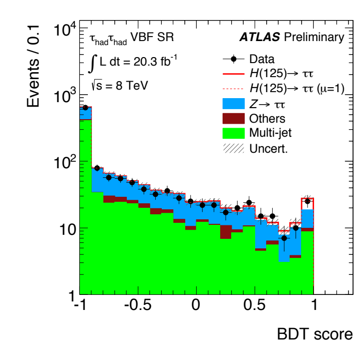
What's a Search?

- Search for New Physics:
  - collect as many events as possible
  - define subsets of events which are likely to be signal, i.e. Signal Regions (SR)
  - estimate amount of background events in SR
  - look at observed data events in SR, estimate compatibility with "Bkg-only" or "Bkg+Signal" hypotheses (test statistic)
- Sensitivity driven by:
  - number of signal events
  - discrimination between signal and background events

Analysis Strategies

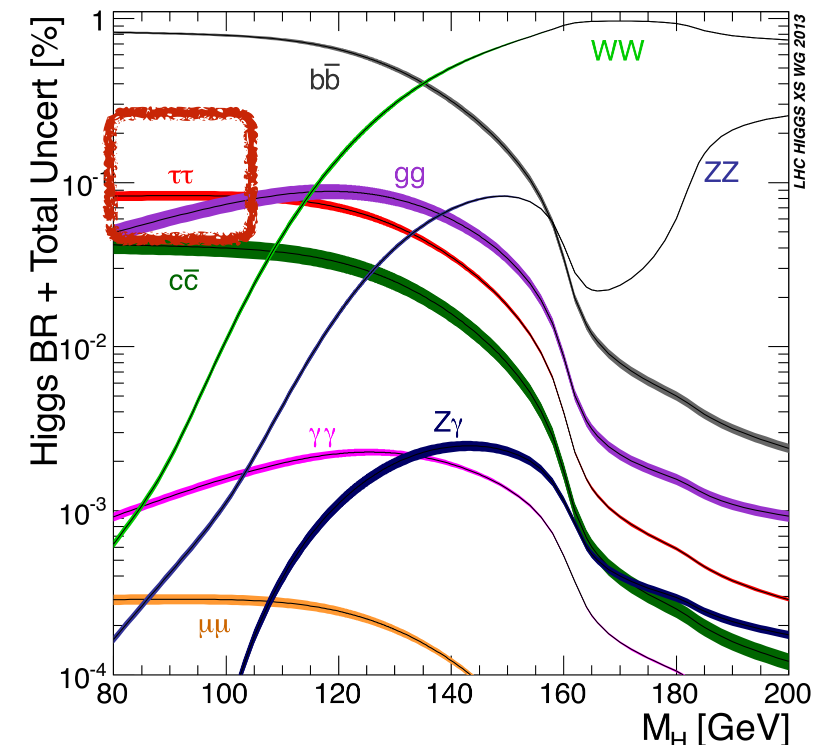
Complexity

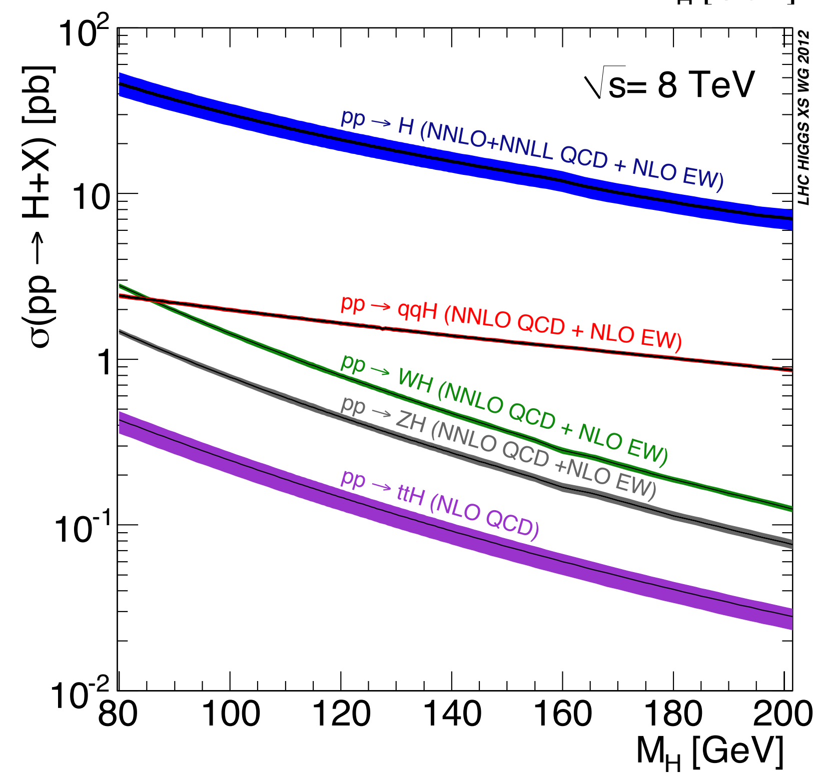
Sensitivity

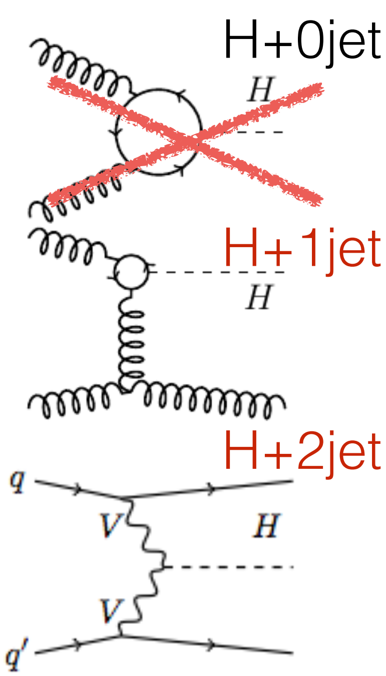
	Signal Region	Statistical Model
Counting Experiment		
“Orthodox” Cut-Based		
Multiple Cut-Based		
Multi-Variate Cut-Based		
Multi-Variate		

Case Study: SM H→$\tau\tau$

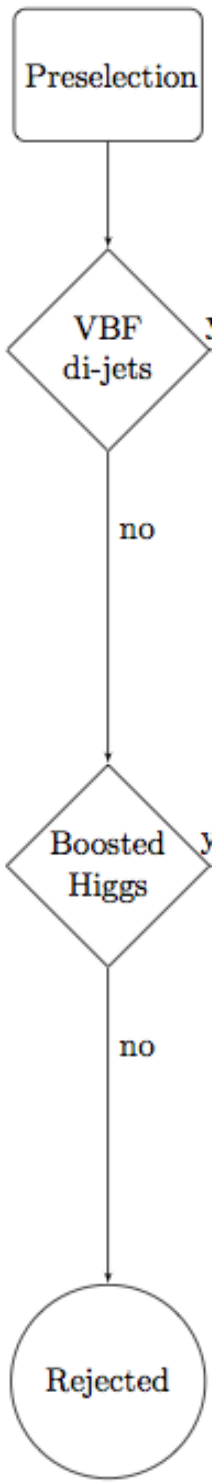
- BR(H→$\tau\tau$)=0.06, 3 final states: ll+4ν (12%), lτ_{had}+3ν (46%), $\tau_{had}\tau_{had}$+2ν (42%)
- Backgrounds: Z→$\tau\tau$ (irreducible), fakes (reducible, i.e. QCD jets selected as leptons or τ_{had}), W+jets, top
- Background rejection:
  - Online (trigger) and offline (object reconstruction, identification) → Strong impact on signal event selection efficiency!
  - Event selection targeting production modes with "tagging" features (ggF H+1j, VBF)
- Invariant mass reconstruction not possible, approximated estimate of neutrino energies → low mass resolution (~15-20%)
- 2 strategies: Orthodox Cut-Based and Multi-Variate Analysis







2 Strategies

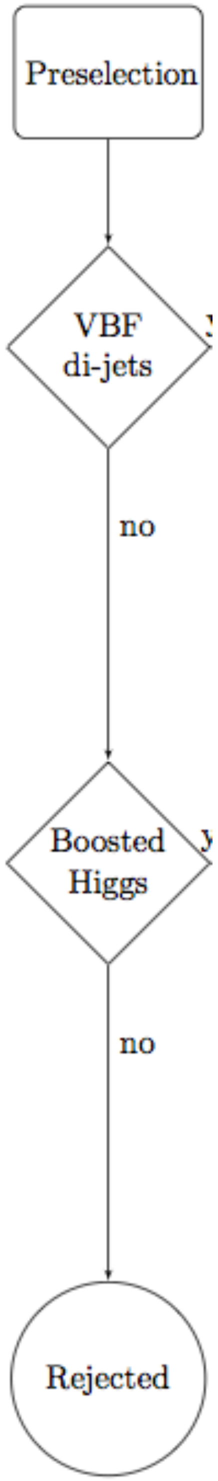


Cut-Based	MVA
Two hard and well separated jets (p_T>50,30 GeV, $\Delta\eta$(j,j)>2.0)	
Cuts: ❖ $\Delta\eta$(j,j)>2.6 ❖ M(j,j)>350 GeV	MVA inputs: ❖ $\Delta\eta$(j,j), M(j,j), η×η ❖ $\Delta R(\tau,\tau)$, **M(τ,τ)** ❖ …
Not accepted in VBF SR, Higgs p_T>100 GeV	
Cuts: ❖ Higgs p_T>140 GeV ❖ $\Delta R(\tau,\tau)$<2.0	MVA inputs: ❖ $\Delta R(\tau,\tau)$, **M(τ,τ)** ❖ τ p_T ratio, Higgs p_T ❖ …
Fit of M(τ,τ)	Fit of MVA output
Strong background rejection, little signal, fit of a physics variable	Loose cuts, many signal events, discrimination by MVA

Comparison: Sensitivity

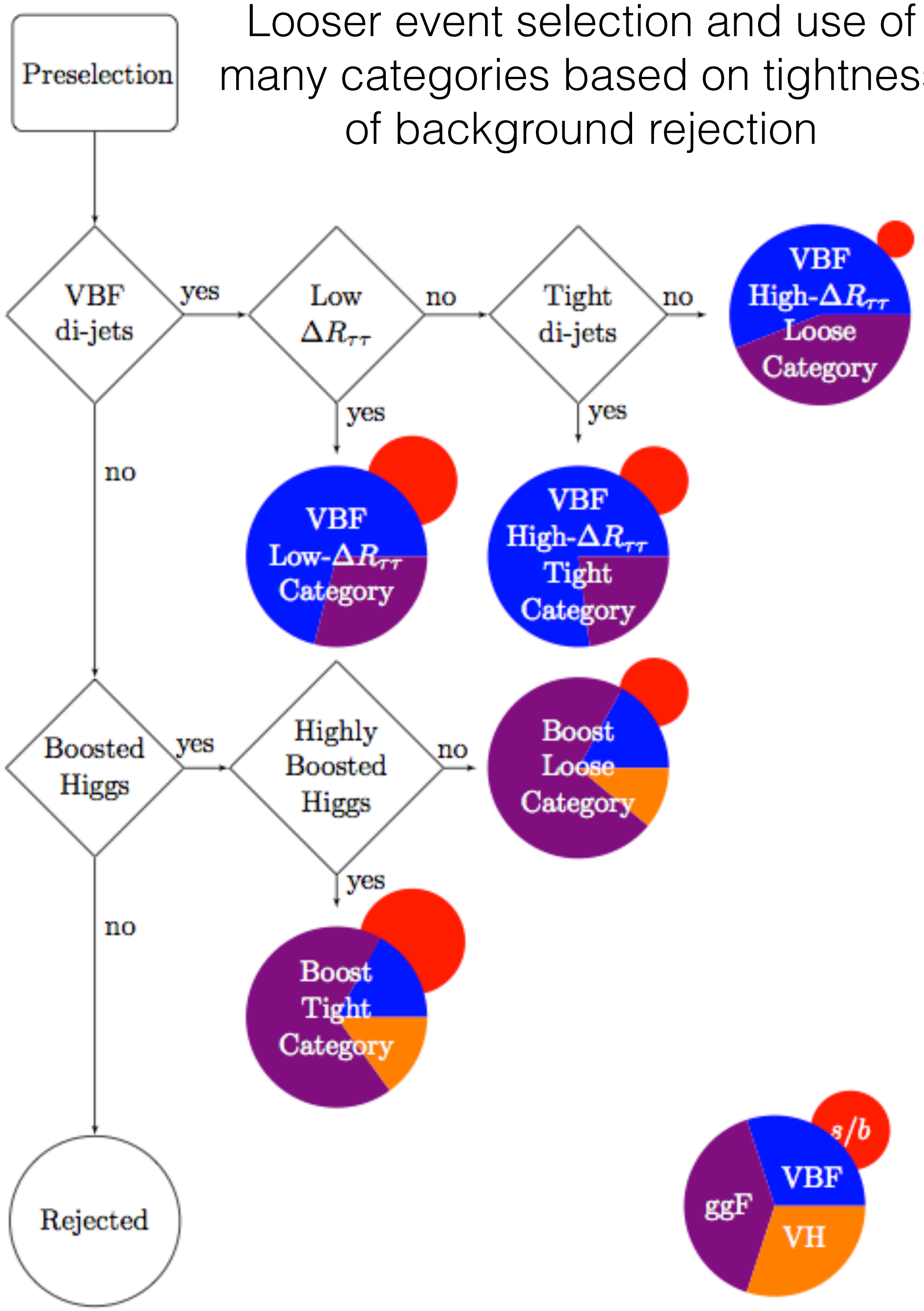
- Significance of signal excess:
  - Cut-Based: 2.3σ
  - MVA: 3.5σ
- Where does this difference comes from? Can we improve the Cut-Based sensitivity?
- MVA profits from:
  - Correlations among event variables, better signal-to-background separation
  - Looser event selection, higher signal acceptance

Multiple Cut-Based



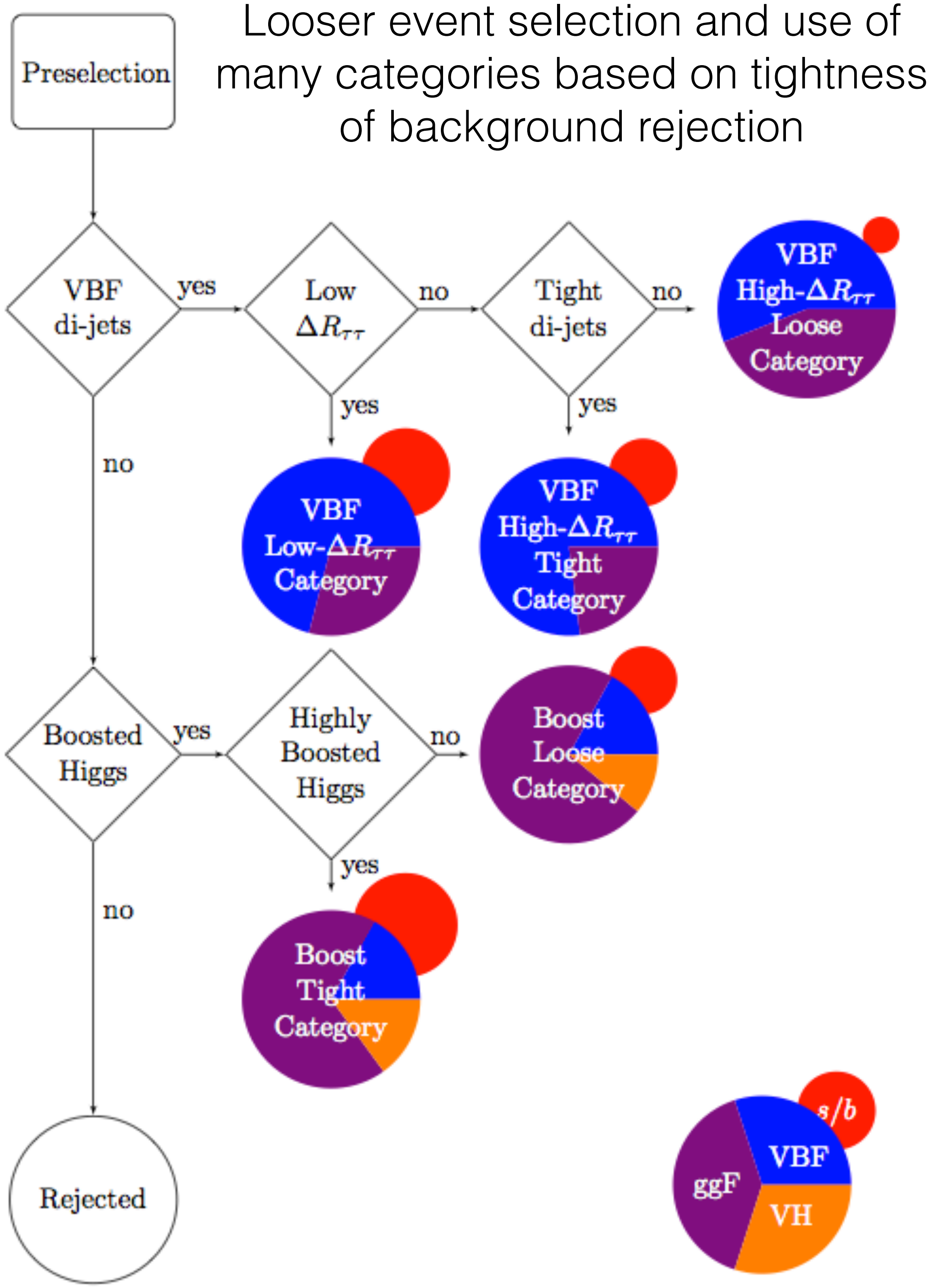
Multiple Cut-Based

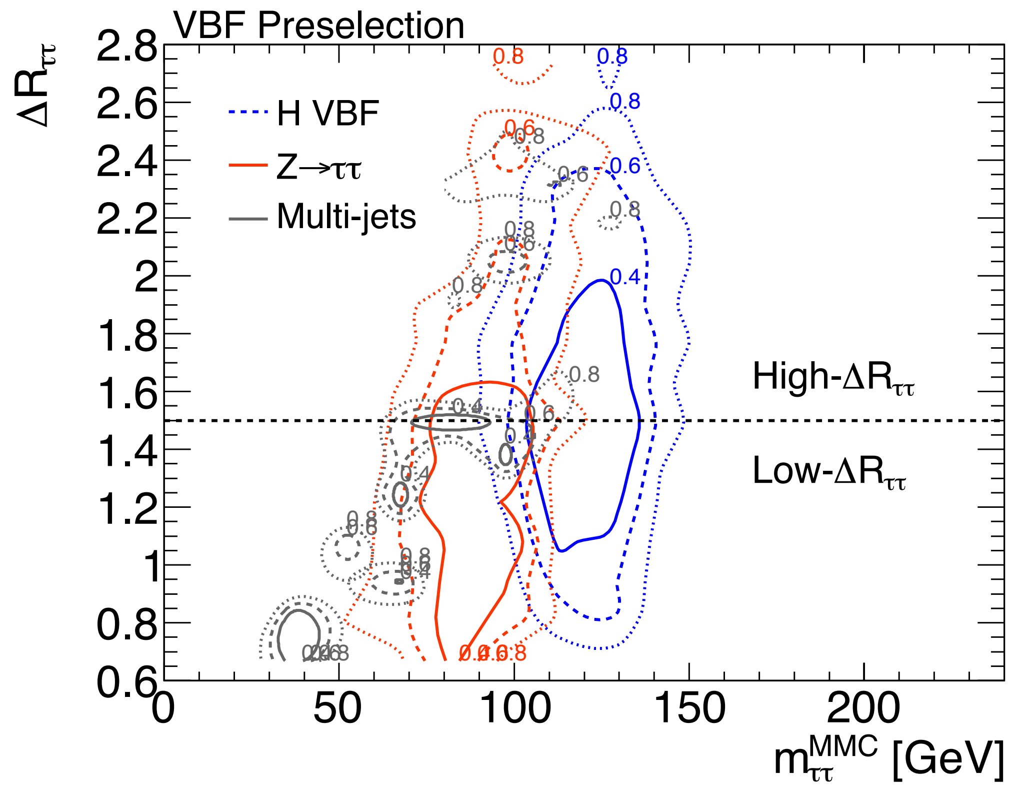
Looser event selection and use of many categories based on tightness of background rejection

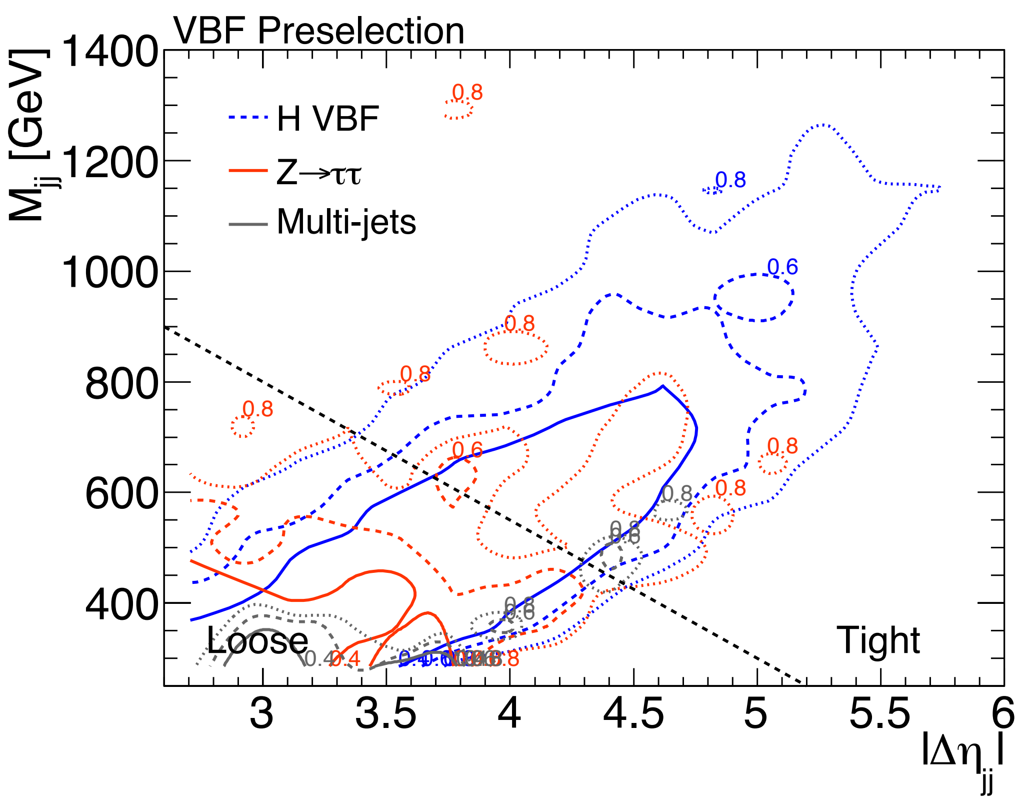


Multiple Cut-Based

Looser event selection and use of many categories based on tightness of background rejection

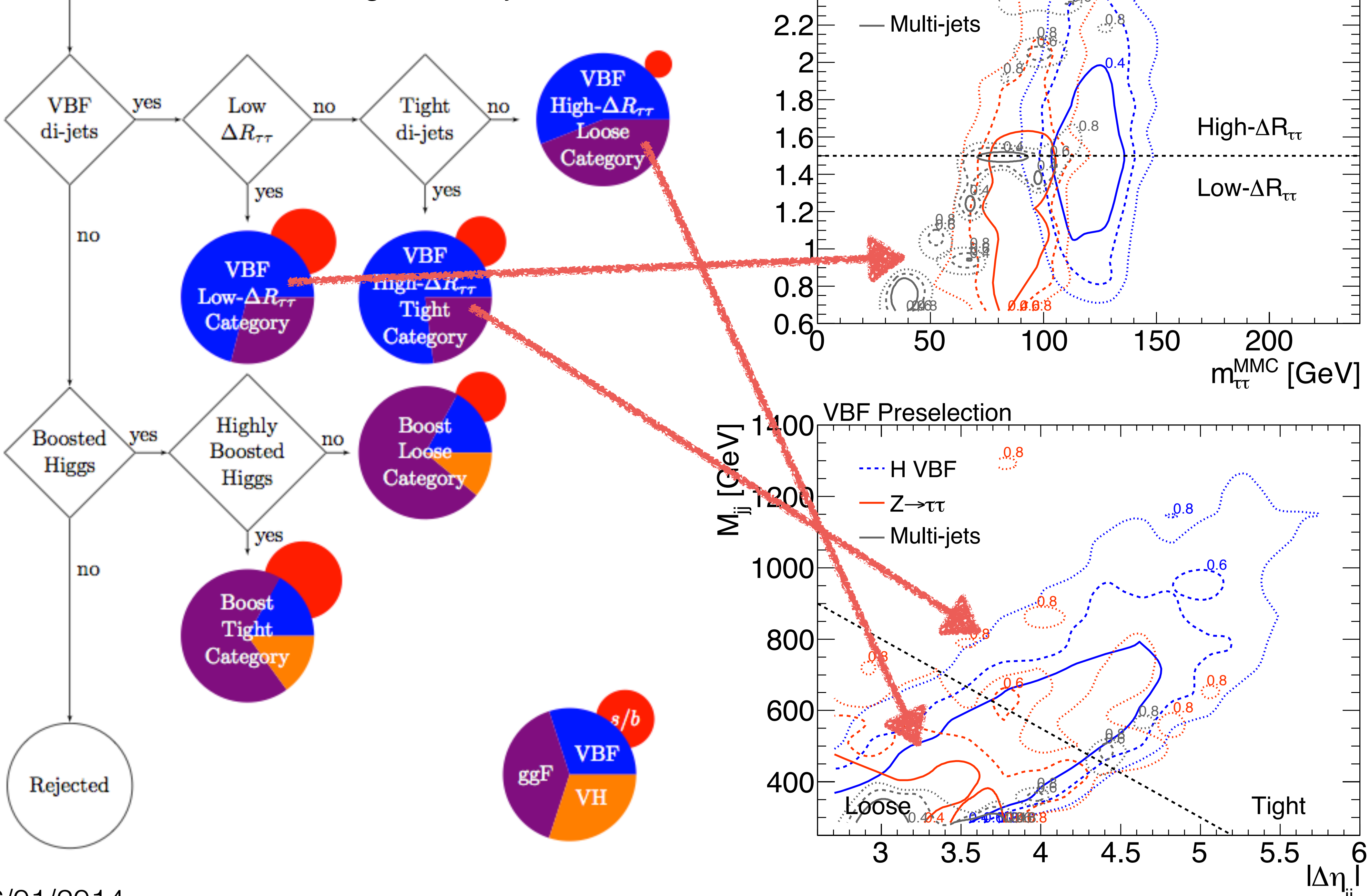






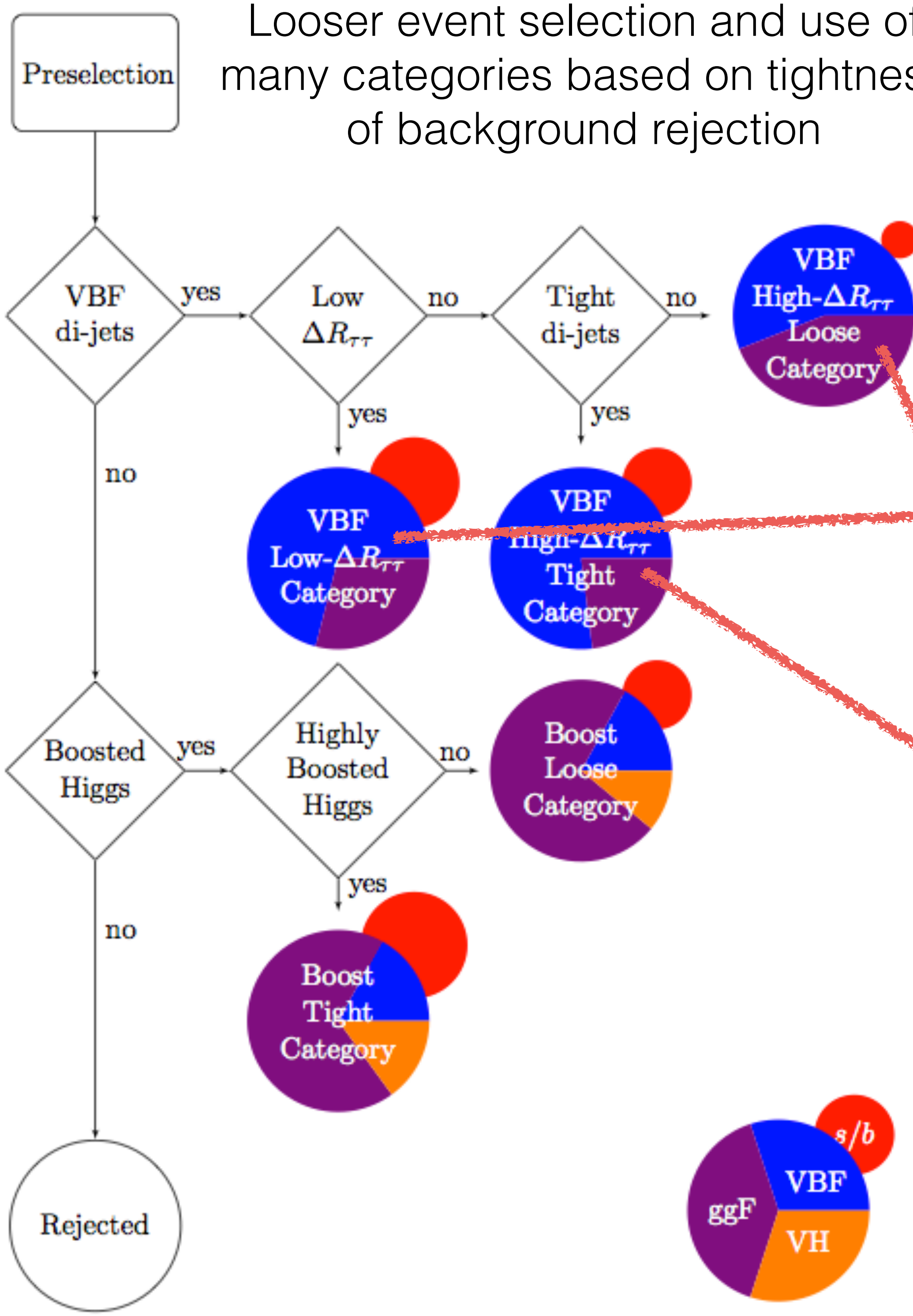
Multiple Cut-Based

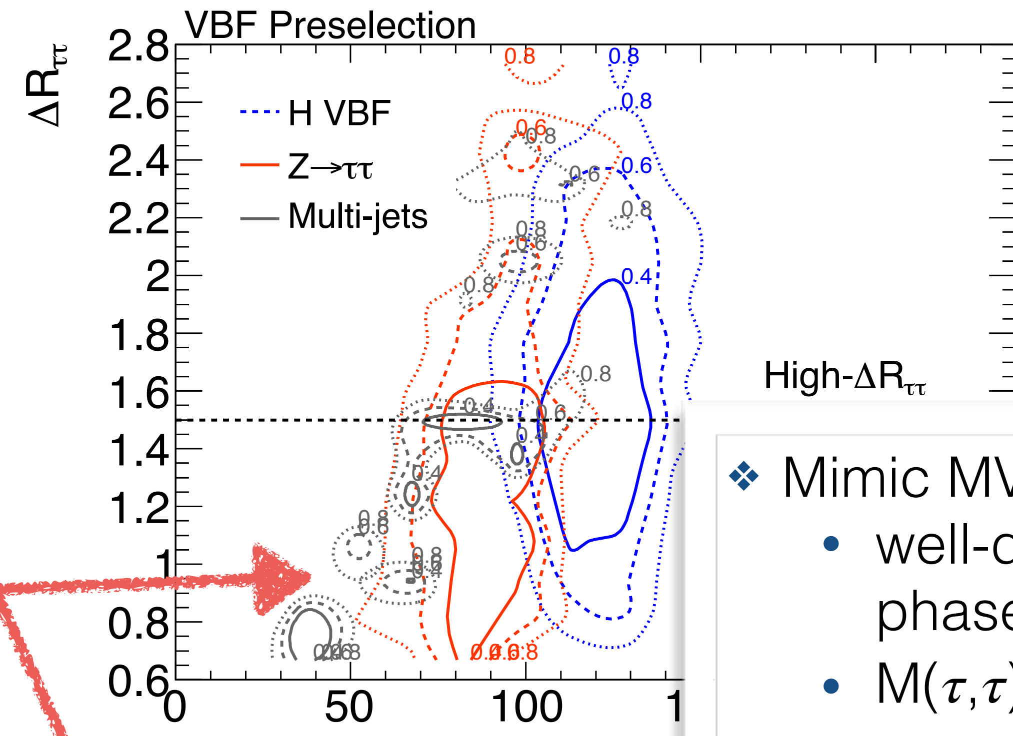
Looser event selection and use of many categories based on tightness of background rejection

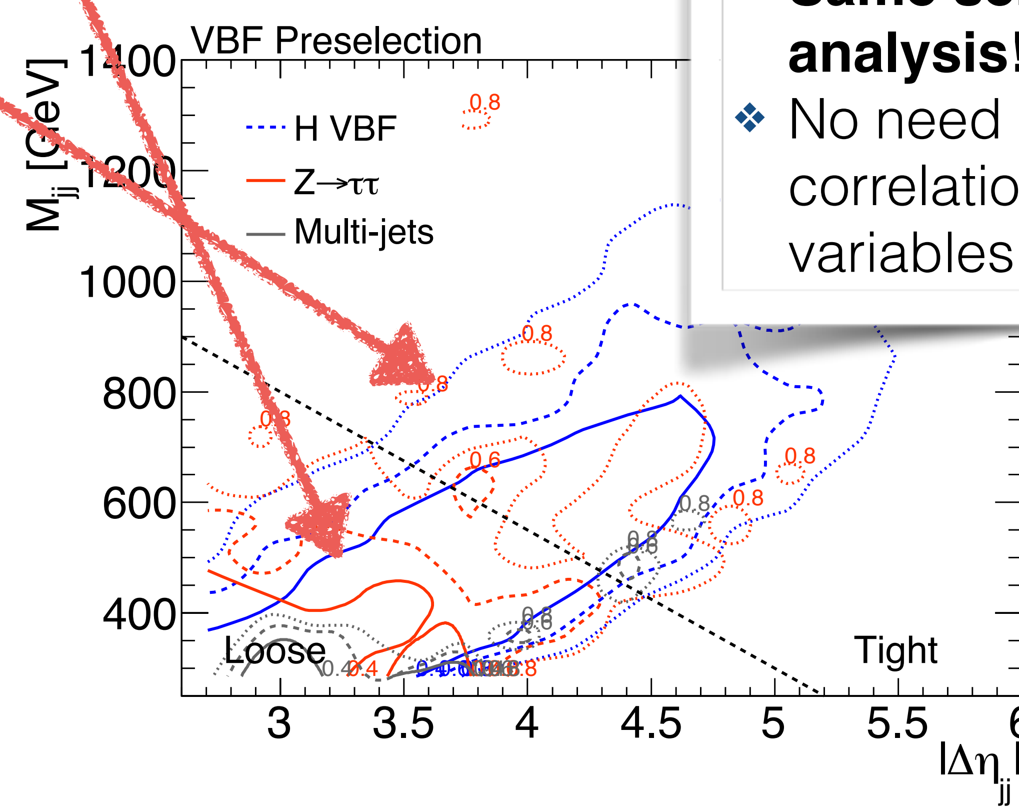


Multiple Cut-Based

Looser event selection and use of many categories based on tightness of background rejection







- Mimic MVA, but with:
  - well-defined kinematic phase space
  - $M(\tau,\tau)$ fit
- **Same sensitivity as MVA analysis!!**
- No need to use all correlations among variables

Comparison: Systematic Uncertainties

	Kinematic Phase Space (Theory Systematics)	Fit (Experimental Systematics)
Multiple Cut-Based	❖ Well-defined categories ❖ Theory systematics: ‣ easy to estimate via QCD scale variations ‣ impacts only on the expected signal yield	❖ Fit of a physical variable, **M(τ,τ)** ❖ Impact of experimental systematic (e.g. energy scales) are predictable ❖ Excesses/mis-modelings can be easily spotted and understood
Multi-Variate	❖ Not well-define signal region, e.g. in VBF SR continuous transition from ggF+≥2jets to ggF+only 2jets ❖ Possible enhancement of **dangerous** corners with poor theoretical predictions ❖ Theory systematics: ‣ no prescription available yet ‣ impacts not only the signal yields but also the **shape** of the MVA score distribution	❖ Fit of a **MVA score** ❖ Impact of experimental systematic cannot be predicted, not trivial to understand ❖ Signal excesses and background mis-modelings more difficult to discriminate

Summary

	Analysis	Motivation
Counting Experiment	❖ SUSY searches	❖ Very tight rejection of SM events ❖ Low statistics
"Orthodox" Cut-Based	❖ H→WW ❖ **H→$\tau\tau$ (old)**	❖ Strong impact of theory and experimental systematics
Multiple Cut-Based	❖ H→$\gamma\gamma$ ❖ H→bb ❖ **H→$\tau\tau$ (CMS)**	❖ Multiple categories to separate classes of events with different signal/bkg ratio
Multi-Variate Cut-Based	❖ VBF H→$\gamma\gamma$	❖ Strong background rejection ❖ Low signal yield
Multi-Variate	❖ **H→$\tau\tau$ (new)**	❖ Need for the best sensitivity to claim 3σ evidence

- Many possible strategies for searches
- Orthodox Cut-Based is too conservative
- Multi-Variate analyses give high sensitivities (best signal-background discrimination)
- Result of Multi-Variate Analyses may be hard to understand (see theory and exp. systematics)
- Advanced Cut-Based analyses can get sensitivities similar to MVA, but with result easier to understand physics-wise

Analysis Strategies

<table>
<tr><th></th><th>Signal Region</th><th>Statistical Model</th></tr>
<tr><td>Counting Experiment</td><td rowspan="2">❖ Subset defined by sequential cuts on events variables
❖ One SR per type of signal (eg SUSY benchmark models or Higgs production modes)
‣ Tight selection: high background rejection, low signal acceptance</td><td>❖ Count of number of observed and expected events in each SR
‣ Simple and robust</td></tr>
<tr><td>“Orthodox” Cut-Based</td><td rowspan="3">❖ Fit of a discriminating variable (eg invariant mass) in each SR
‣ Result “easy” to interpret physics-wise (eg mass peak)</td></tr>
<tr><td>Multiple Cut-Based</td><td>❖ (Cut-Based)^N: multiple SRs per type of signal
‣ Potential use of correlations among event variables, higher signal acceptance</td></tr>
<tr><td>Multi-Variate Cut-Based</td><td rowspan="2">❖ Subset defined by the output of a MVA algorithm (BDT, Neural Net., …)
❖ Multiple event variables used as input for MVA
‣ Use of correlations among event variables, strong background rejection AND good signal acceptance</td></tr>
<tr><td>Multi-Variate</td><td>❖ Fit of the output of the MVA algorithm
‣ Best search sensitivity
‣ Result not easy to interpret</td></tr>
</table>