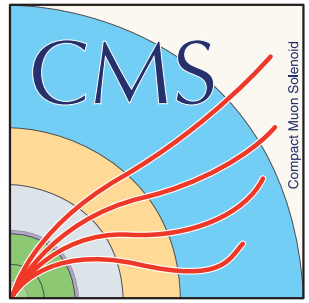
Imperial College London



Imperial College Status

Albert Dow
on behalf of IC

Higgs CP in τ Decays Workshop
DESY, Hamburg
10 Oct 2019

a.dow16@imperial.ac.uk

Overview

I. Brief general status update

II. Latest expected sensitivity

III. Signal vs. background MVA studies

IV. Legacy 2016

V. Synchronisation

VI. Plan

General Status

Focus on all-hadronic $\tau\tau$ channel with decay plane method.

Expected sensitivity of (non-optimised) analysis has been extracted with 80X 2016 set-up.

Developed and using MVA for increasing τ decay mode efficiencies and purities.

Current effort is to get the legacy 2016 (94X) set up, together with new production versions of 2017 and 2018.

Additions are the vertex refit, inclusion of deepTauv2p1 and puppiMET, new JEC for 2018 data/MC.

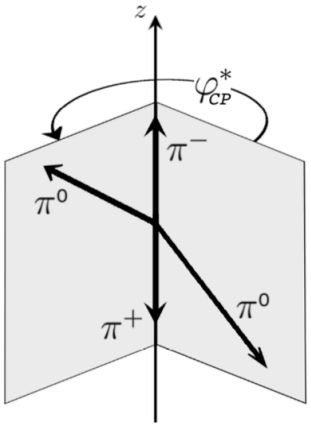
Fake factors (FF) for specific for the CP in decays analysis.

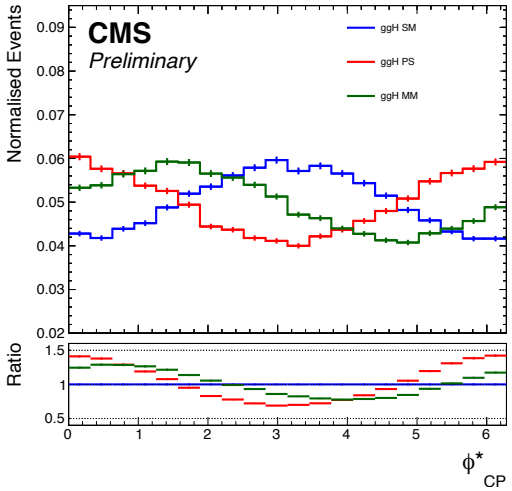
Decay Plane Method

Working mainly on the all-hadronic τ channel with decay mode $\tau \rightarrow \rho$ (+ν), where $\rho \rightarrow \pi\pi^0$, with τ branching fraction of 25.5%.

Also combined (so far) with final state $\tau\tau \rightarrow a_1\ \rho$ (+2ν) contributes 4.6% of $H \rightarrow \tau\tau$ decays.

Work in the CoM frame of the two charged prongs. Strip (π^0) can suffer from resolution issues at RECO level.





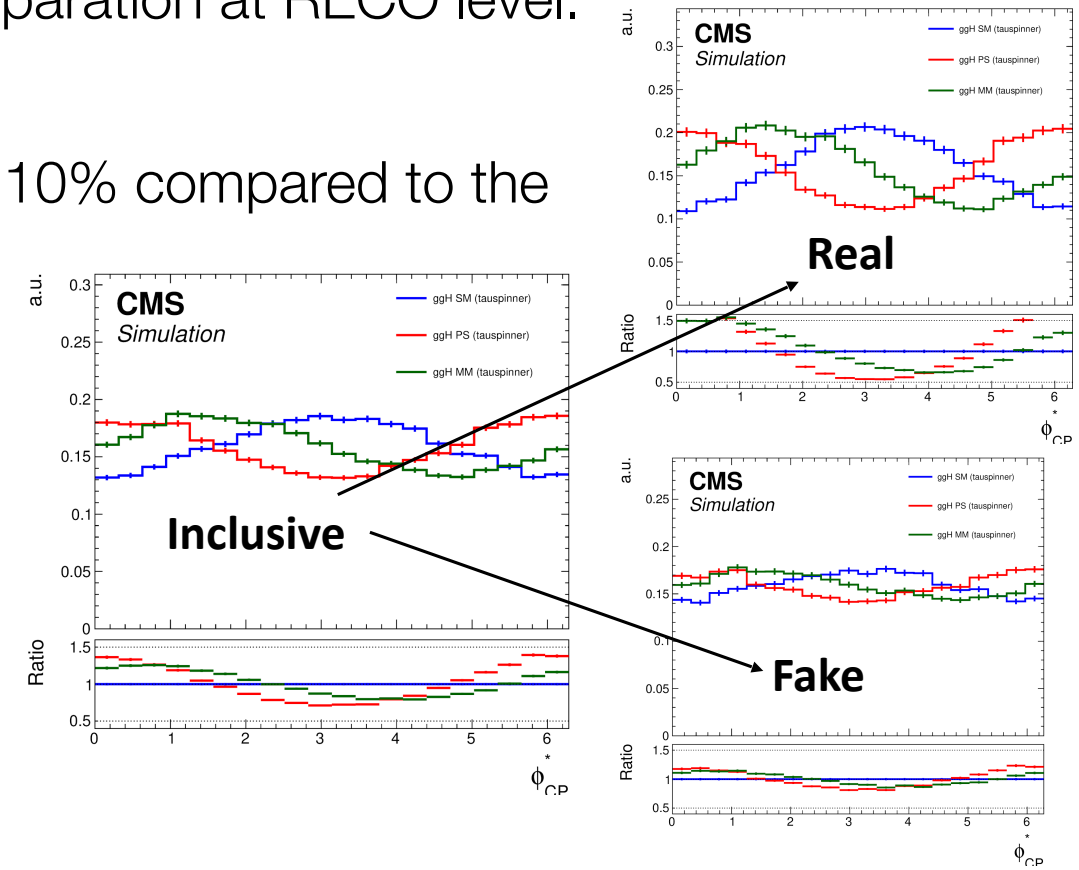
Decay Mode MVA

Developed a specific MVA in order to separate “real" and “fake” ρ decays, as these smear out the separation at RECO level. (see figures on right).

Gain in purity and efficiency about 10% compared to the RECO-based τ decay mode.

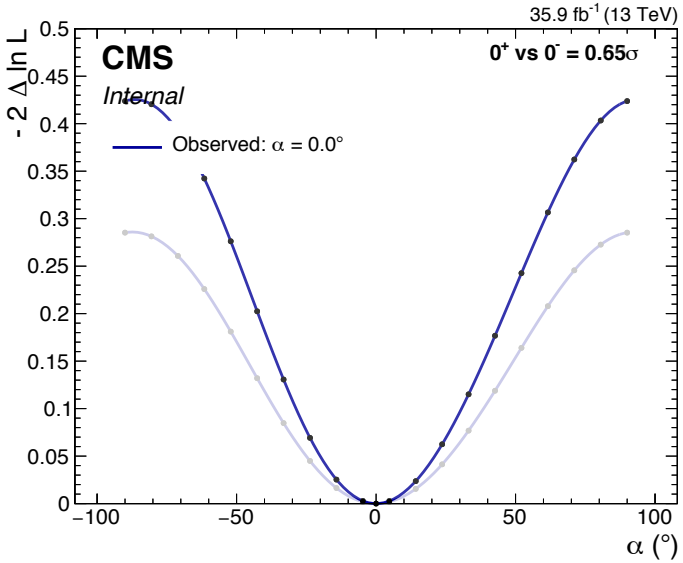
Similar improvements seen for the MVA to enhance the 3-prong a_1 purity in RECO decay mode = 10.

(See Mohammad’s talk)



Gain in Sensitivity from MVA DM

Previously showed the gain in (expected) sensitivity for $\rho\rho$ and $a_1\rho$ combined with the MVA DM gains by about 22% compared to without MVA DM.



Background Shape Checks

Idea is to check the shape of the CP angle distribution in the $\rho\rho$ decay (HPS DM).

In data also select ZTT and QCD/SS separately, using OS data and m_sv < 100 GeV (ZTT-enriched), and SS data requirement for QCD-enriched. For MC, just used stacked histograms as usual.

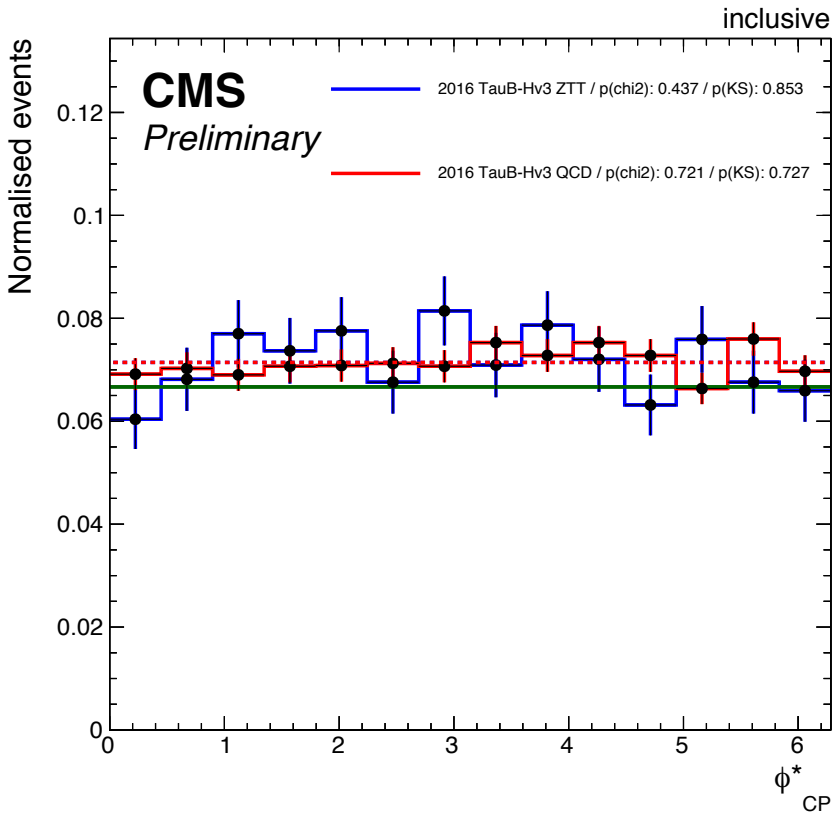
Plotting Data (on left) and MC (on right) to avoid showing unblinded data on top of MC.

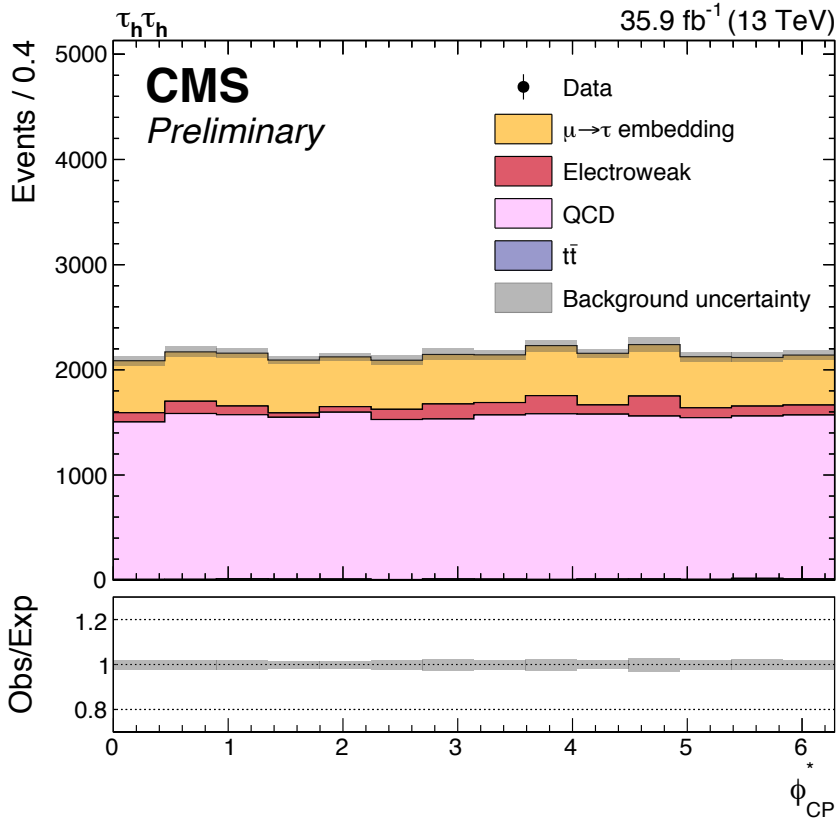
Then fitting to the background- enriched data shapes using the likelihood method, also writing out the associated p-value. Reject the hypothesis if $p < 0.05$.

Included a flat (green) histogram with which I am doing the KS goodness of fit test.

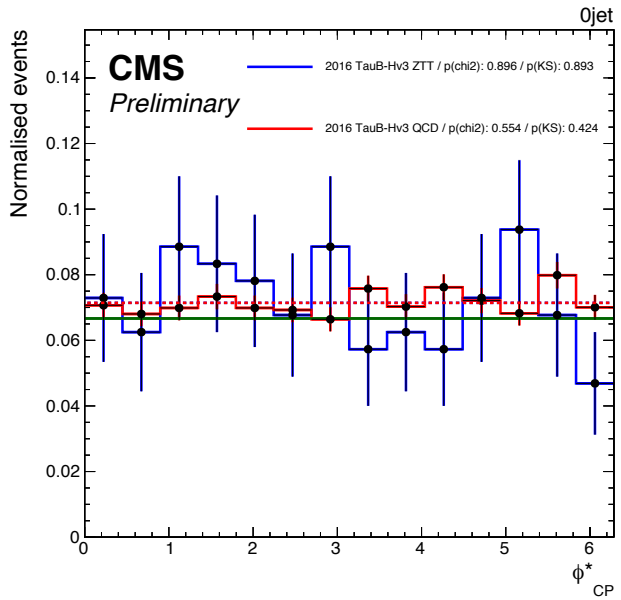
Overall, backgrounds look flat in both data and MC. Can merge bins to gain bin-by-bin statistics.

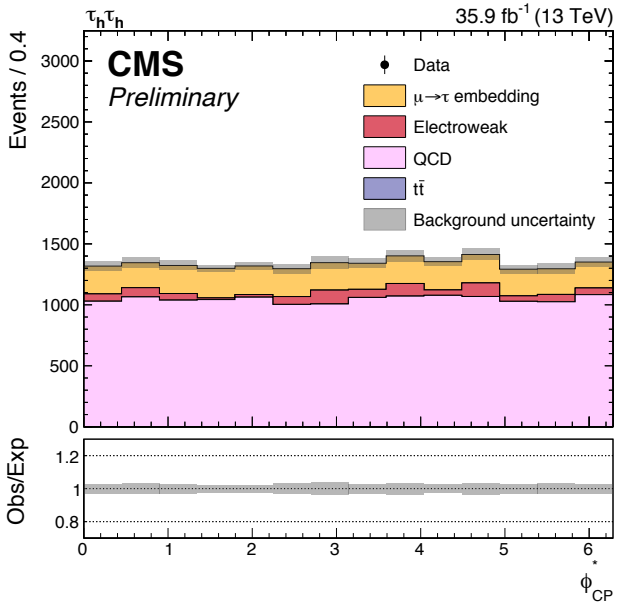
inclusive


inclusive
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.437 / p(KS): 0.853
2016 TauB-Hv3 QCD / p(chi2): 0.721 / p(KS): 0.727
Normalised events
0.12
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
4
5
6
ϕ^*_{CP}

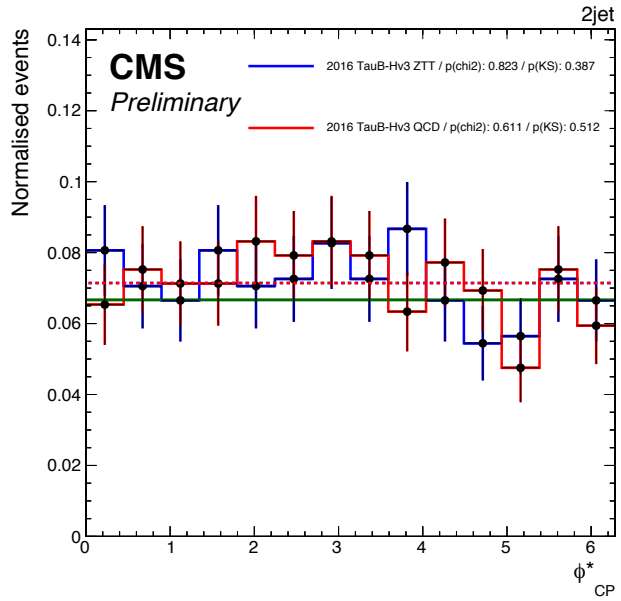

$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
5000
4000
3000
2000
1000
0
Obs/Exp
1.2
1
0.8
0
1
2
3
4
5
6
ϕ^*_{CP}

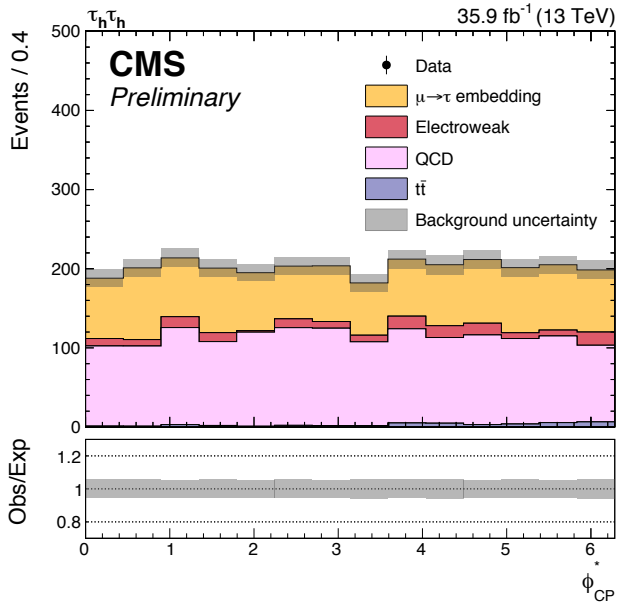
0jet


0jet
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.896 / p(KS): 0.893
2016 TauB-Hv3 QCD / p(chi2): 0.554 / p(KS): 0.424
Normalised events
ϕ^*_{CP}


$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
Obs/Exp
ϕ^*_{CP}

dijet


2jet
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.823 / p(KS): 0.387
2016 TauB-Hv3 QCD / p(chi2): 0.611 / p(KS): 0.512
Normalised events
ϕ^*_{CP}


$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
Obs/Exp
ϕ^*_{CP}

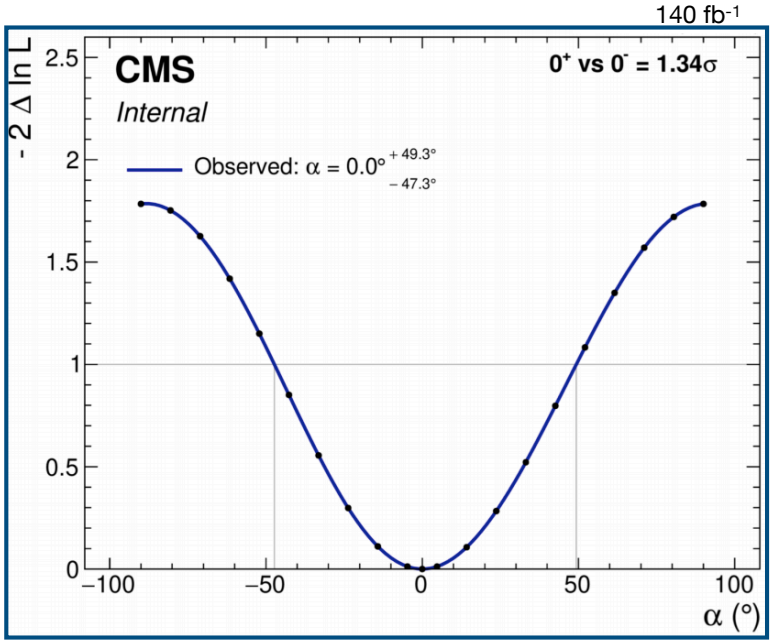
Expected Sensitivity

Based on 80X samples produced with no spin effects and reweighted using tauspinner.

MVA is trained on embedded samples and using the FF method → mainly data-driven for $\tau\tau$ channel.

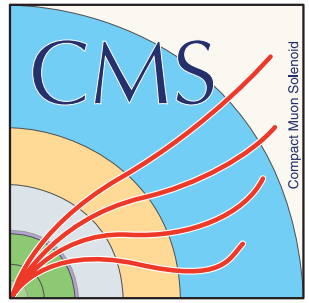
To be improved by inclusion of deepTauID and a_1 a_1 channel.

Uses $\rho\rho$ and $a_1\rho$ channels together with MVA DM for increasing purity of ρ and a_1.



Additionally, can merge background bins as can see that the background is flat in both data and MC.

Imperial College London



MVA Studies

I. Sensitivity differences when training on SM or PS

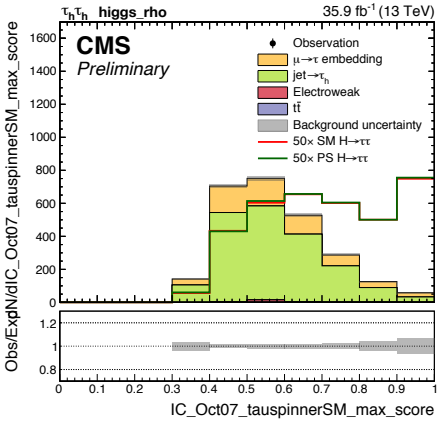
II. Train with 1 (“higgs”) or 2 signal classes (“ggH”, “qqH”)

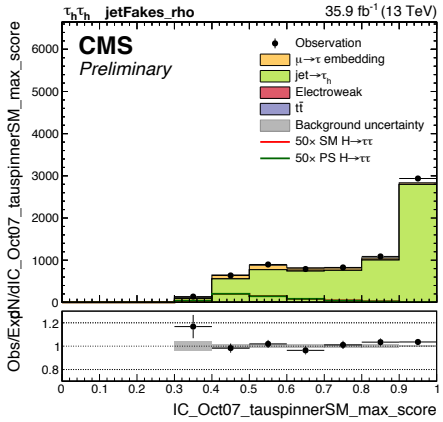
Train with SM vs PS

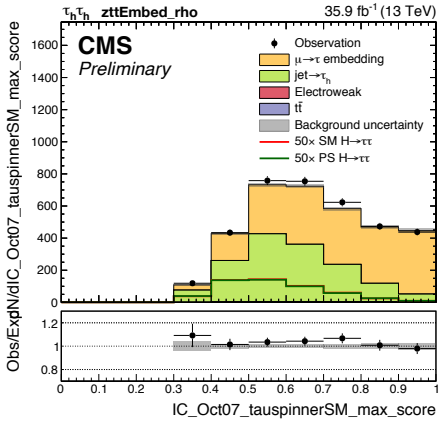
Checked the effect of training (BDT) on tauspinner SM and PS samples on the fit.

Switched to training on tauspinner (with genfilter) means a lot more statistics to previous iterations done with powheg.

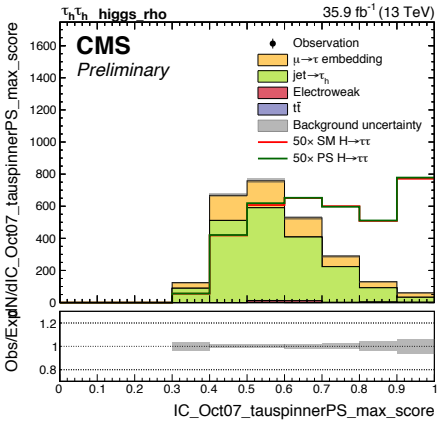
SM

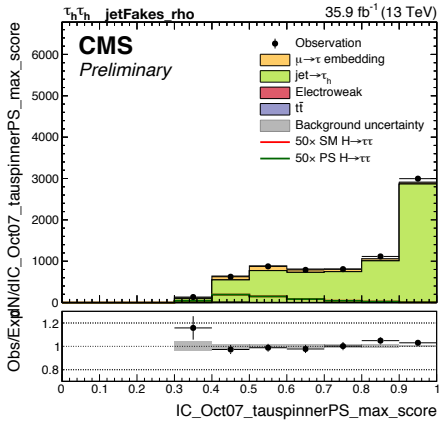


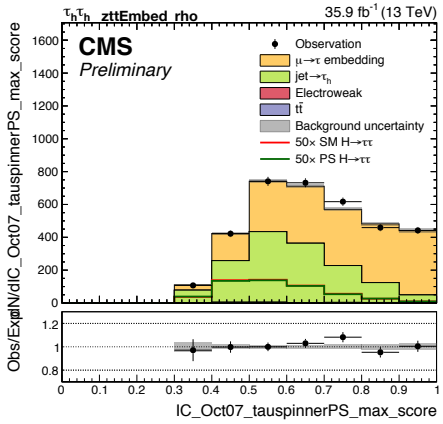




PS



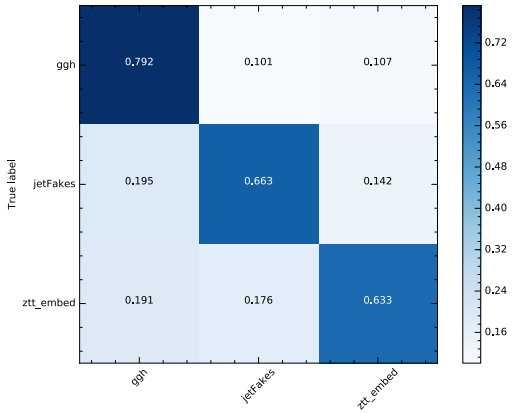


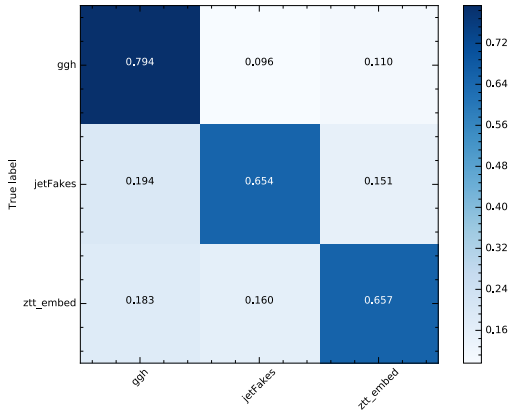


Train with SM vs PS

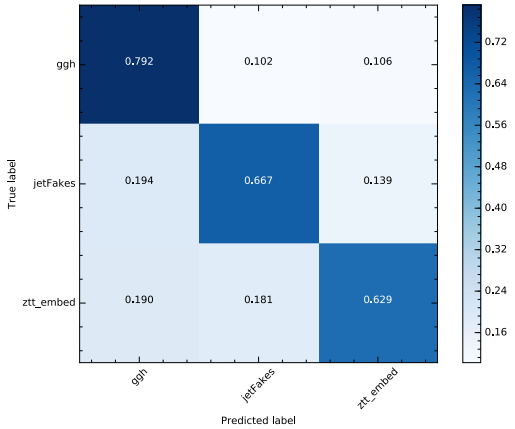
On first order can check the confusion matrices for the classification efficiency — not much difference.

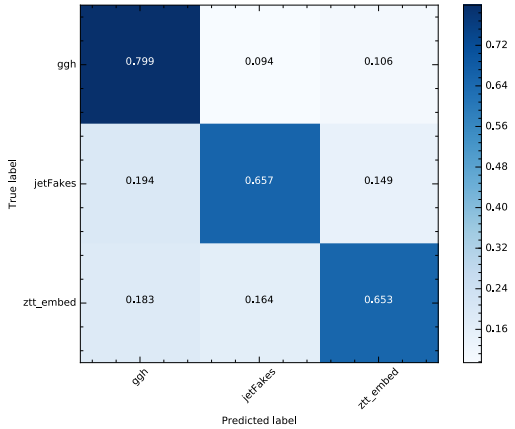
SM





PS





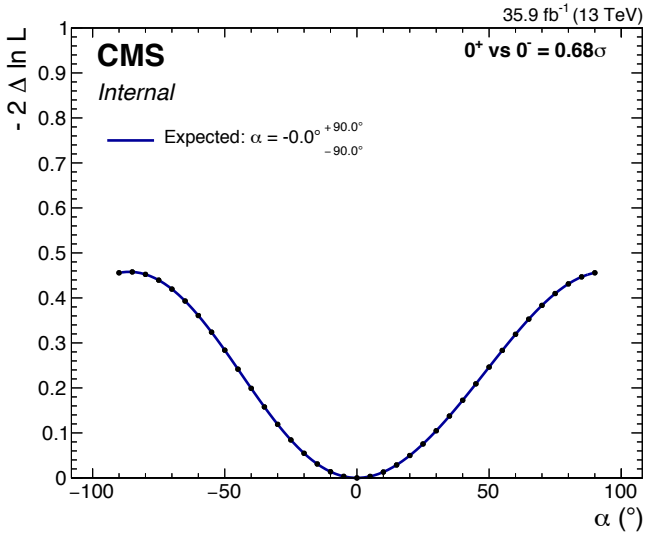
Train with SM vs PS

Very similar results on fit level as well.

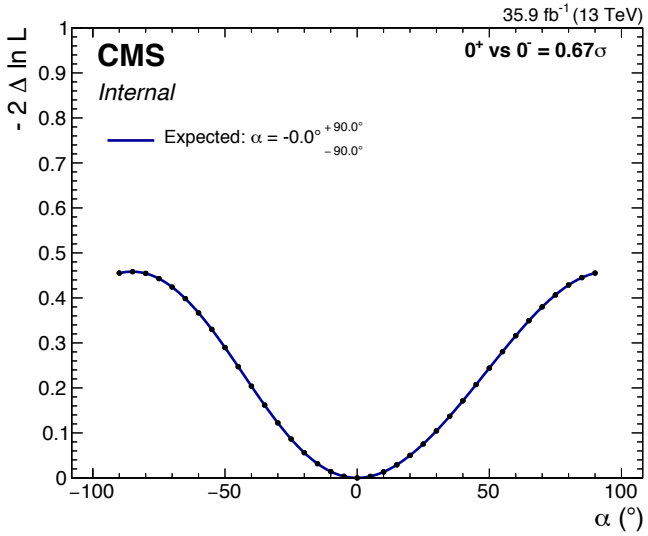
Fit done with (full) $\tau\tau$ systematic template.

Could conclude that tauspinner does not bias our result — could be different with a NN?

Additionally similar result from using tauspinner vs. powheg samples.



SM



PS

Train on Higgs Samples

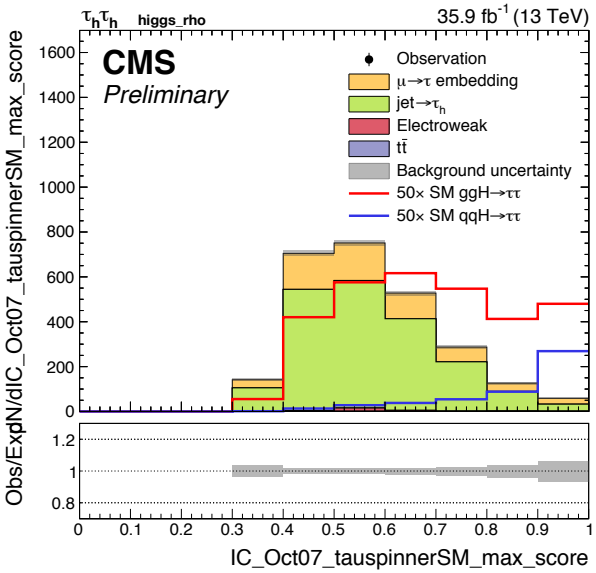
Do we want to train on Higgs (ggH and qqH) as one class or as two separate classes?

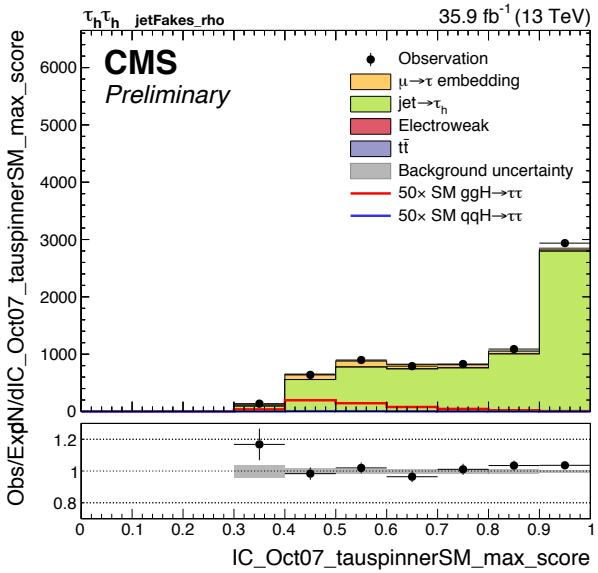
Next two slides show the different distributions for the MVA output categories with separate ggH and qqH signals.

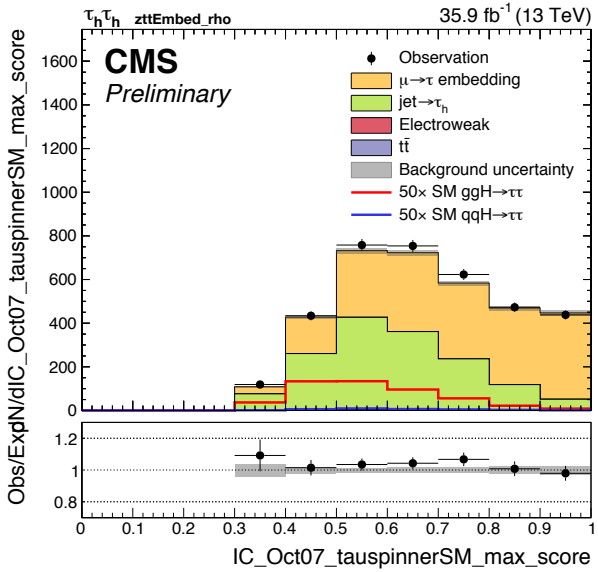
These control plots are all in the $\rho\rho$ decay channel (HPS DM = 1).

Train on ggH and qqH together

Here showing same outputs from training just mentioned in the previous SM vs. PS training study. Red is ggH, blue is qqH.





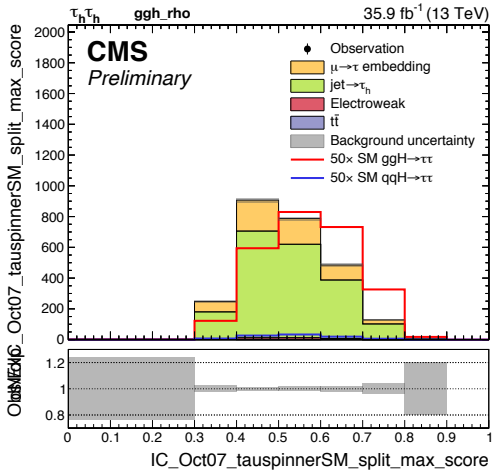


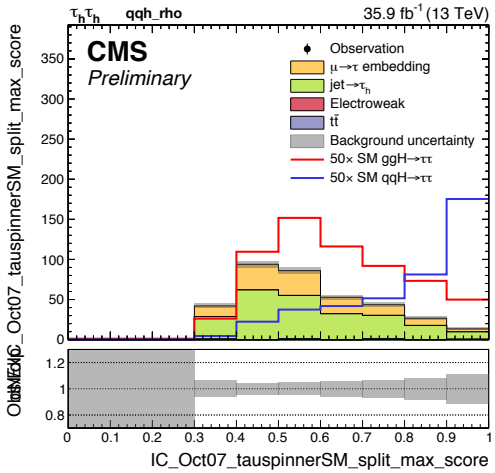
Train on ggH and qqH separately

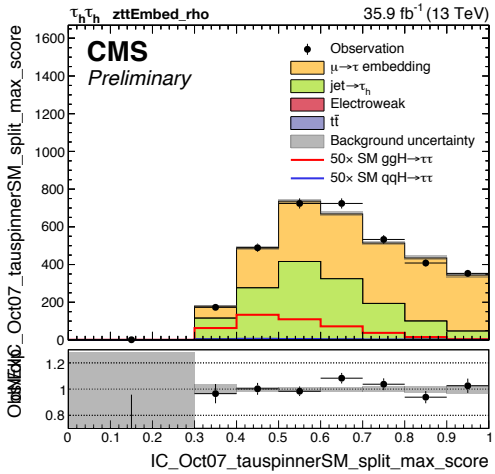
Less ggH is pushed towards the right in this case, compared with previous slide.

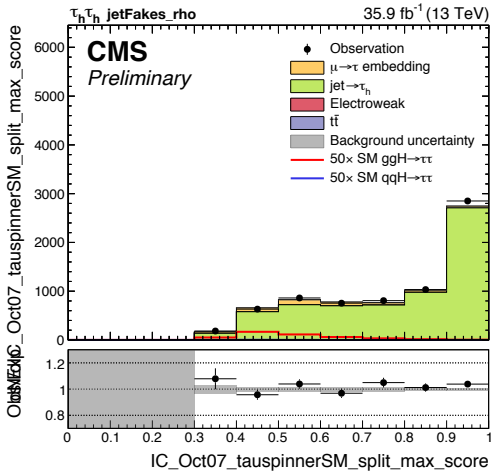
ggH peaks more with Z peak.

Training ggH and qqH inclusively optimises more towards VBF and high dijet mass events — boosted with a more distinct topology.



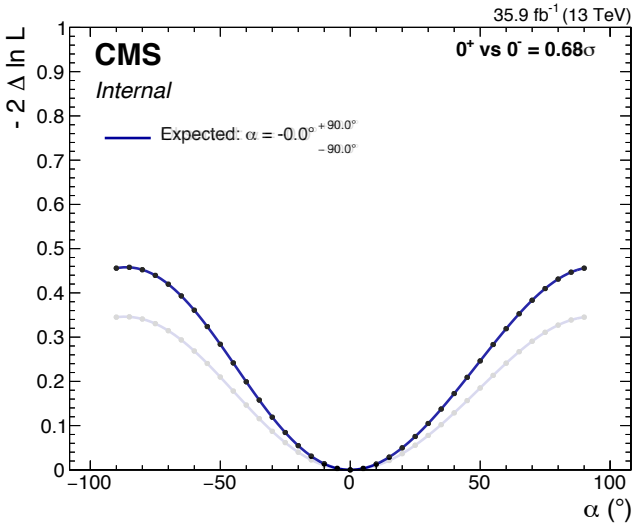


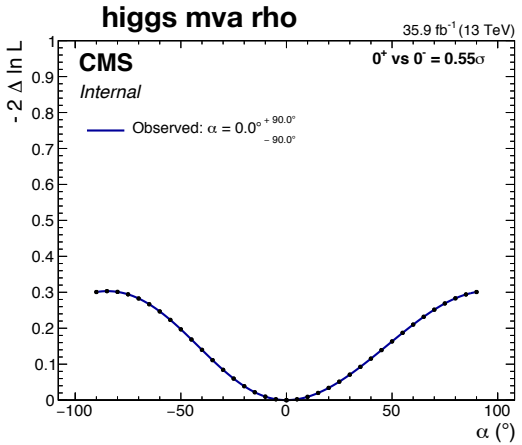


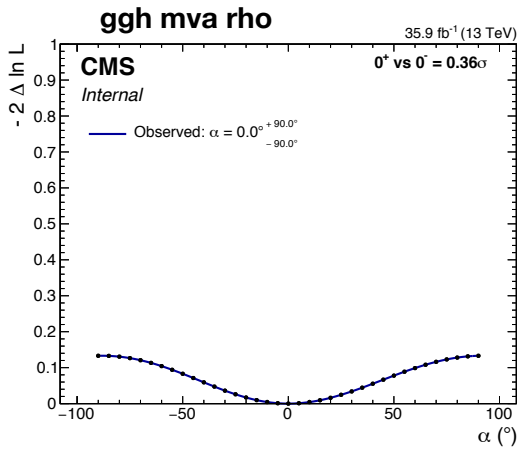


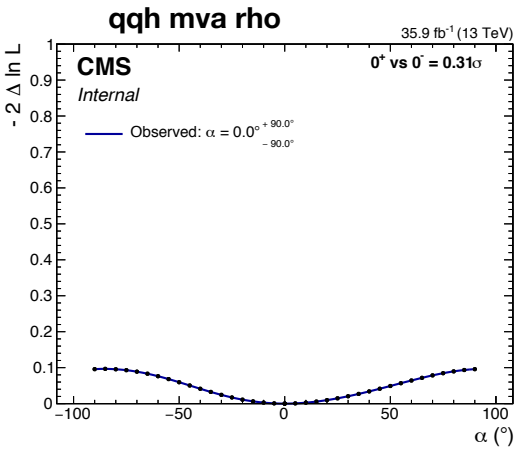
Train on ggH and qqH together

Splitting the higgs training classes is about 15-25% worse in terms of expected sensitivity.





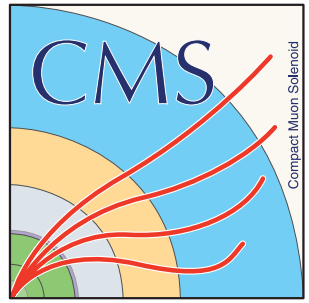




Top right: higgs (inclusive training) MVA $\rho\rho$ channel
Bottom right: ggH and qqH MVA $\rho\rho$ channels.

Less optimised for VBF when splitting H training categories.

Imperial College London



Legacy 2016

Legacy 2016 (94X)

Important switch to Legacy 2016 due to different τ reconstruction — more consistent with 2017 and 2018.

Embedded samples now ready — to be included soon together with required corrections.

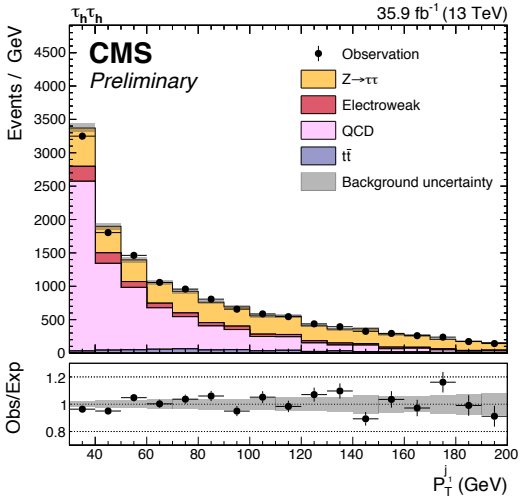
Current set of corrections based on [*], some to be measured by IC for checks.

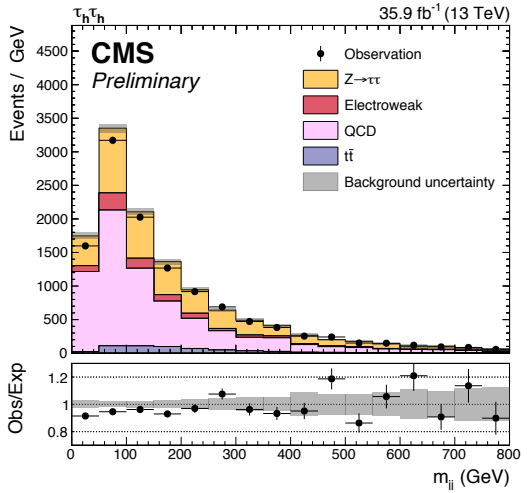
FF measured by Danny already (see his talk).

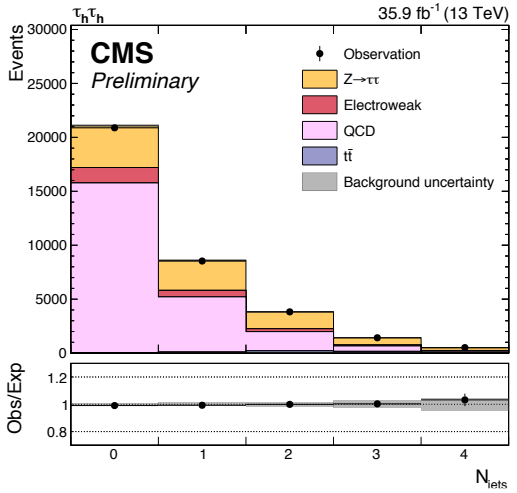
[*] https://github.com/KIT-CMS/LegacyCorrectionsWorkspace

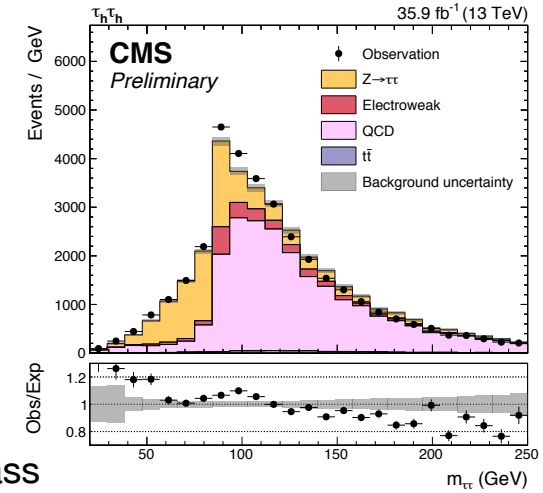
Legacy 2016 (94X)

Jet variables reasonably well described — more work required to describe τ objects better.

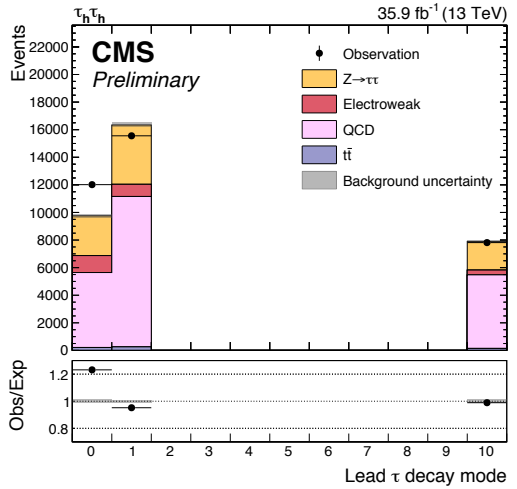








visible mass



Synchronisation

First set of ntuples produced with

2016: /VBFHToTauTau_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM

2017: /VBFHToTauTau_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM

2018: /VBFHToTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v1/MINIAODSIM

On top of standard HTT sync variables, deepTauID has been added, tauspinner weights, MVA tau decay modes.

Additionally switched to puppiMET.

Found in

/afs/cern.ch/work/a/adow/public/Sync/<year>/SYNCFILE_VBFHToTauTau_M-125_<year>_<channel>.root

Needs further iterations to add all components needed as per [*].

[*] https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsCPinTauDecaysSync

Plan

Finish setting up legacy 2016 together with 2017 and 2018 eras.

Currently for 2016 (94X) and 2018 no samples exist, 2017 have (small) signal samples.

> Idea: for 2016 94X — use shapes from 80X samples and scale to what is expected in new samples; for 2018 — use 2017 samples.

Set up CombineHarvester for 2018 — 2016 and 2017 (almost) ready. Systematic template needs updating: correlations, uncertainties on ρ and π position/resolution?

Run a comparison of SVFit versus fastMTT and conclude based on the (expected) signal strength which is better to use — ideally fastMTT because of speed!

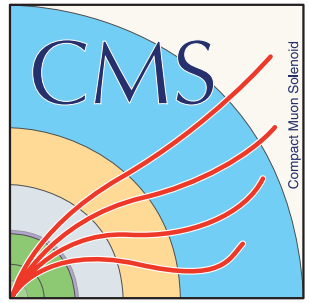
Train signal vs. background BDTs for all 2017 and 2018 as well and combine sensitivities.

Add a_1 a_1 (3pr) channels — possibility of combining all angles into a better observable?

Focus on continuous synchronisation efforts.

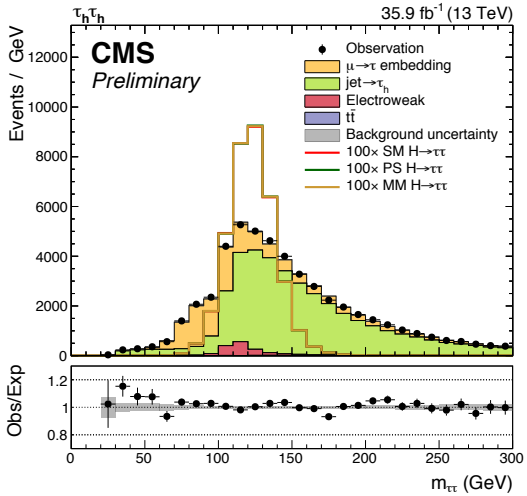
Current full run2 extrapolation yields ~1σ. With final set-up, including deepTau ID, expect at least the same. Ultimately combine categories from other channels!

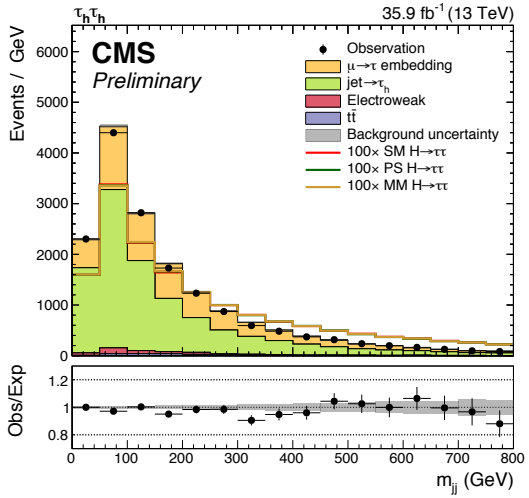
Imperial College London

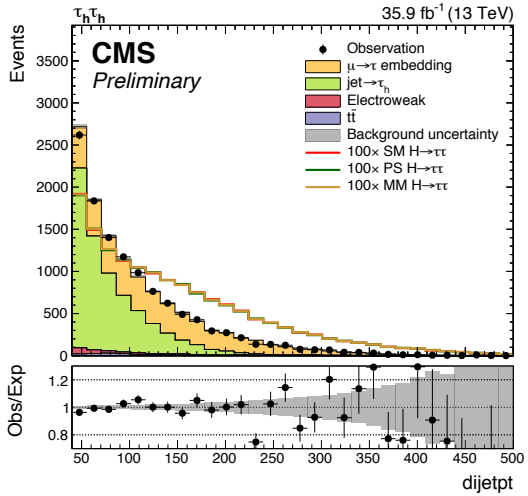


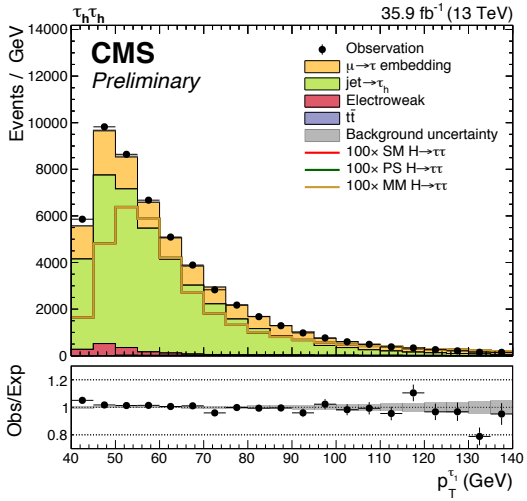
Backup

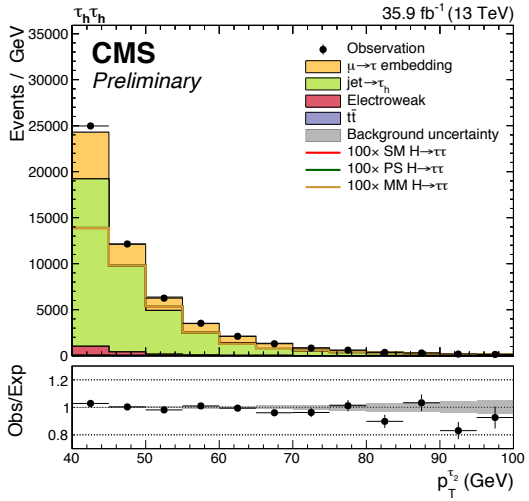
Input variables (inclusive)

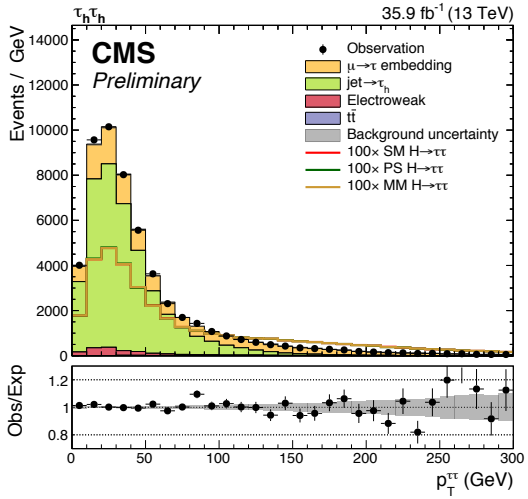






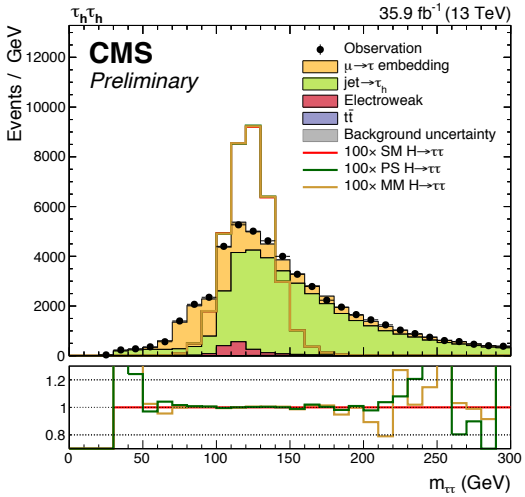


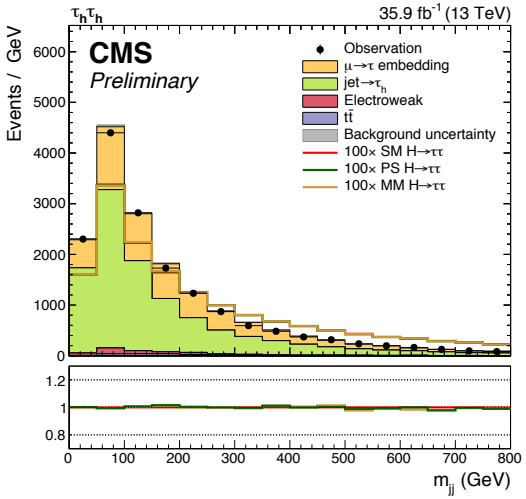


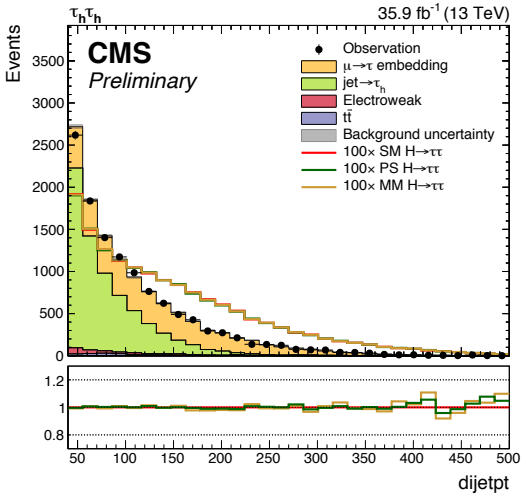


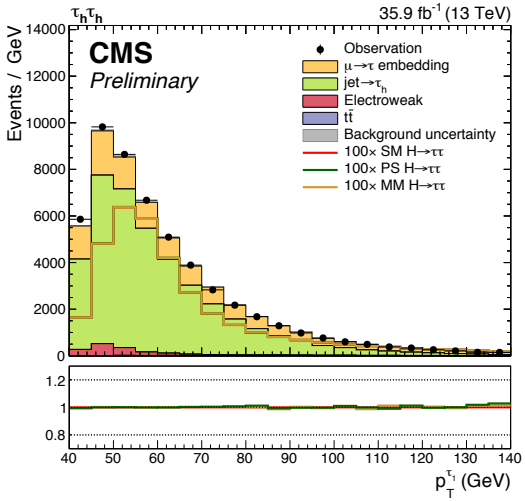
Input variables (inclusive):

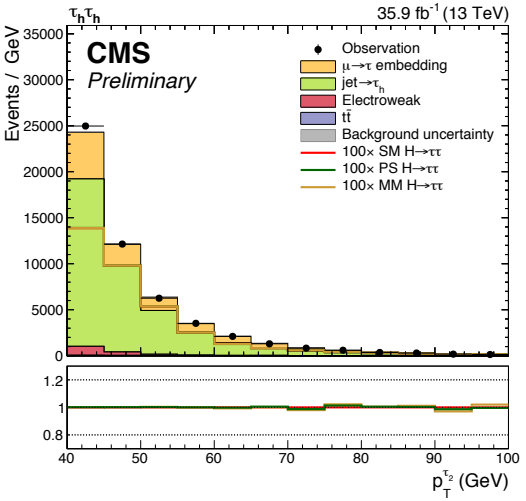
ratio of signals with respect to SM signal

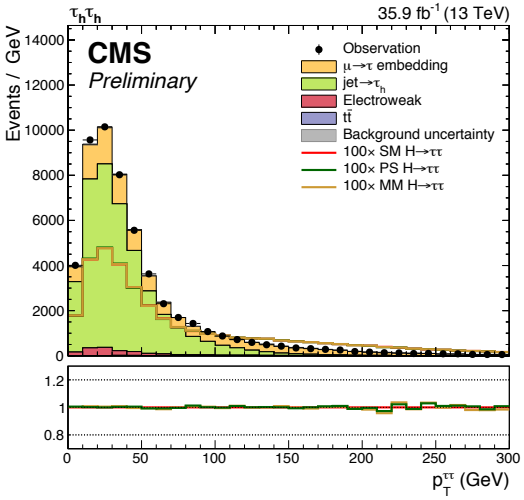






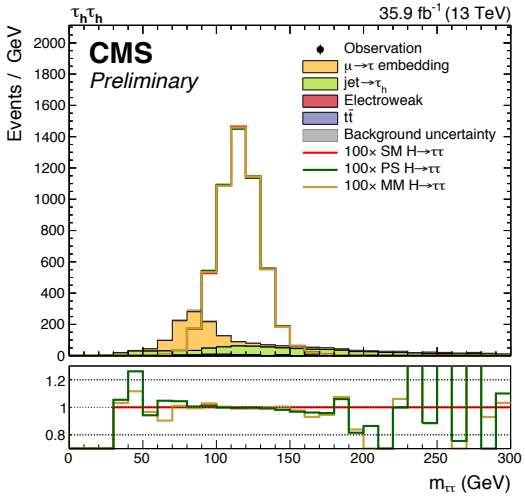


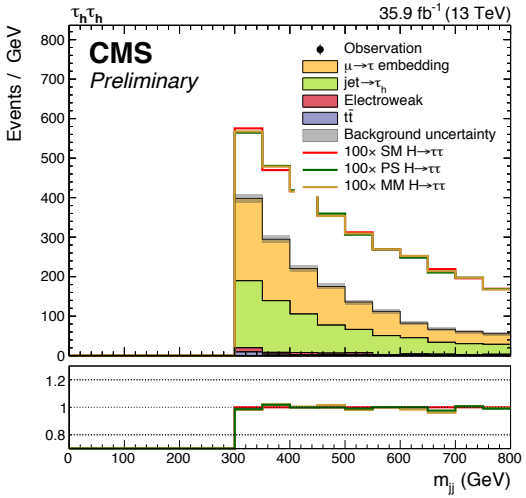


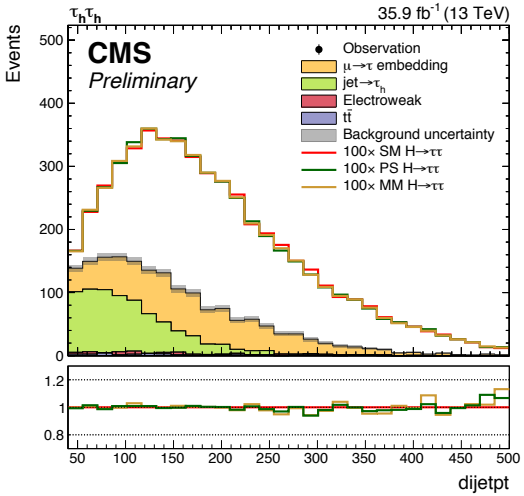


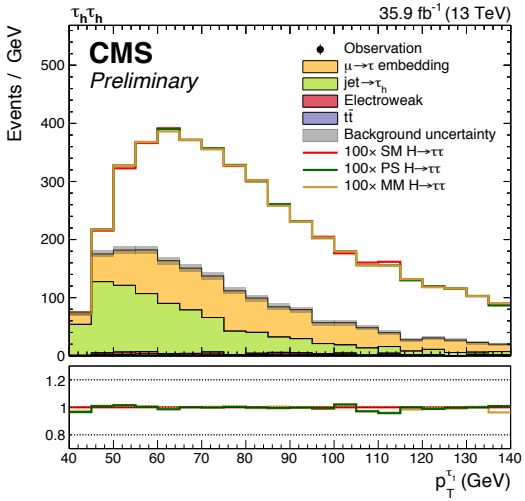
Input variables (dijet + mjj > 300):

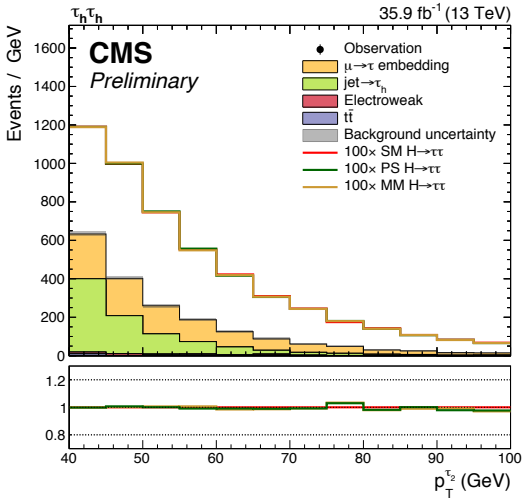
ratio of signals with respect to SM signal

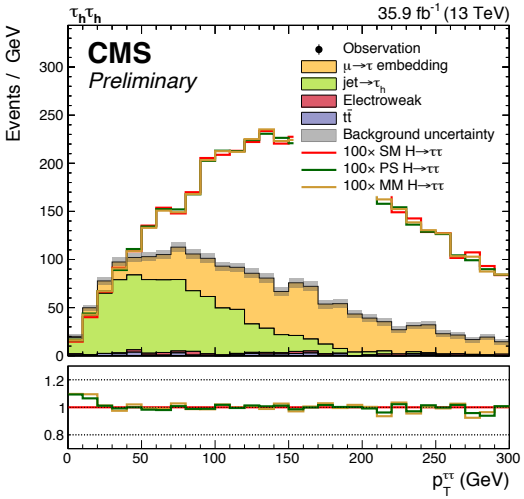










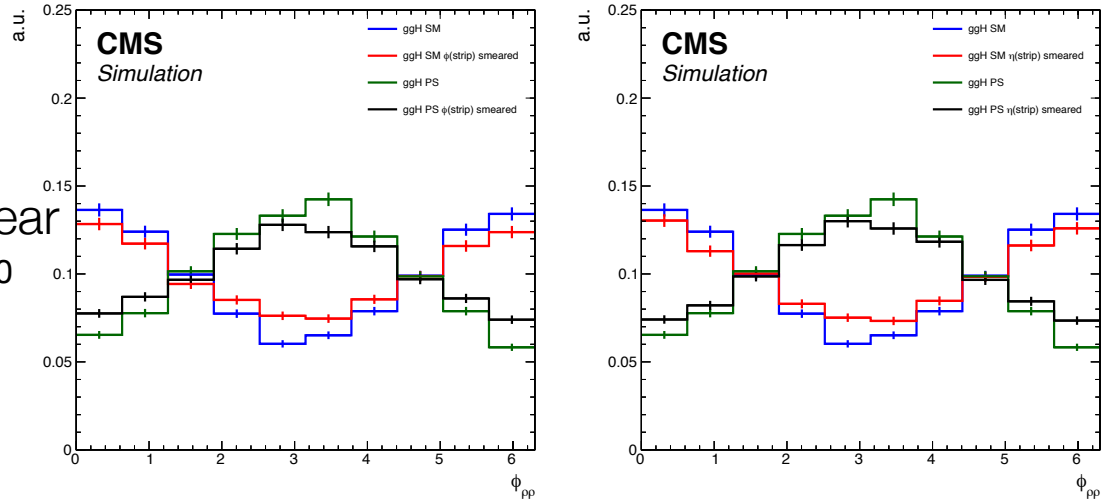


Effect of Smearing on φ_{CP} in $\rho\rho$

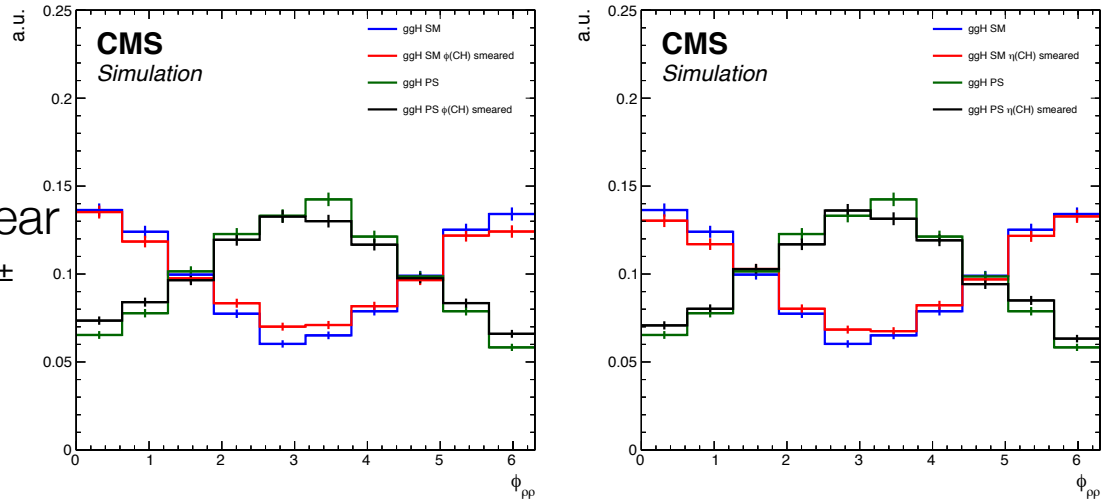
Previously found that smearing the strip's (angular) position results in smaller amplitude on CP distribution in $\rho\rho$ case, especially φ.

Charged π smearing has less of an impact.

Smear π^0



Smear $\pi^{\pm}$



Background Shape Checks

Idea is to check the shape of the CP angle distribution in the $\rho\rho$ decay.

Standard baseline for $\tau\tau$ channel with $\rho\rho$ decays (just HPS DM, no MVA DM).

In data also select ZTT and QCD/SS separately, using OS data and m_sv < 100 GeV (ZTT-enriched), and SS data requirement for QCD-enriched. For MC, just used stacked histograms as usual.

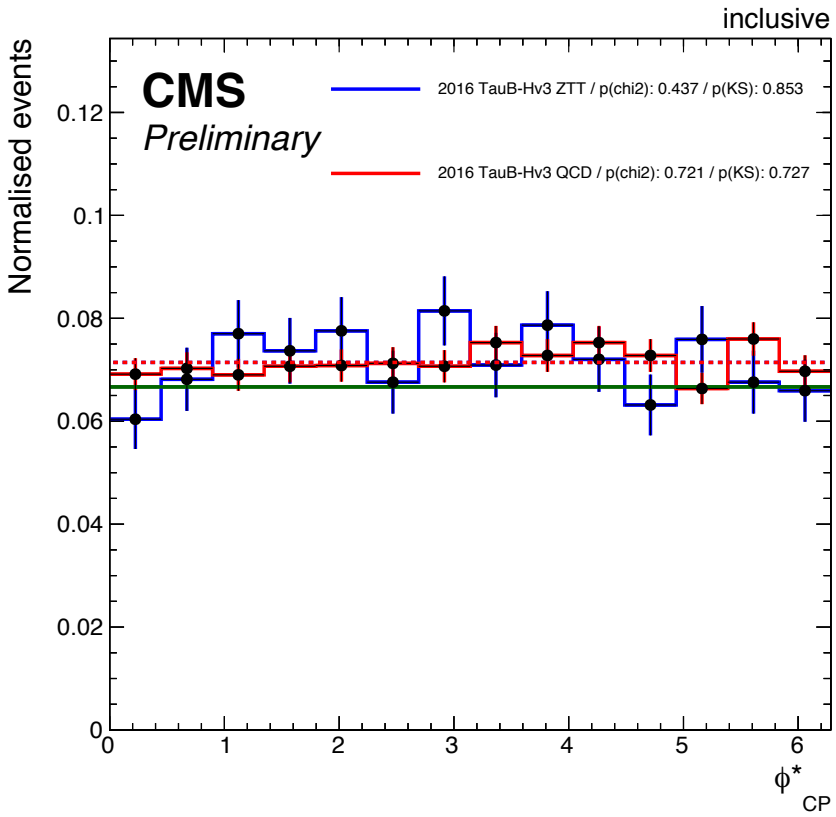
Plotting Data (on left) and MC (on right) to avoid showing unblinded data on top of MC.

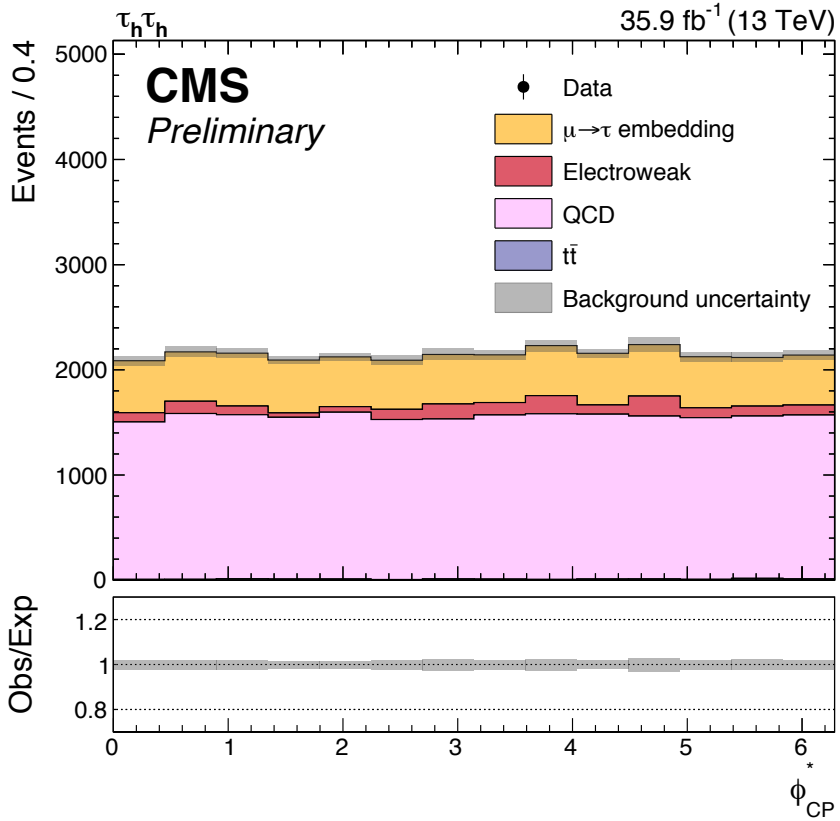
Then fitting to the background- enriched data shapes using the likelihood method, also writing out the associated p-value. Reject the hypothesis if p < 0.05.

Included a flat (green) histogram with which I am doing the KS goodness of fit test.

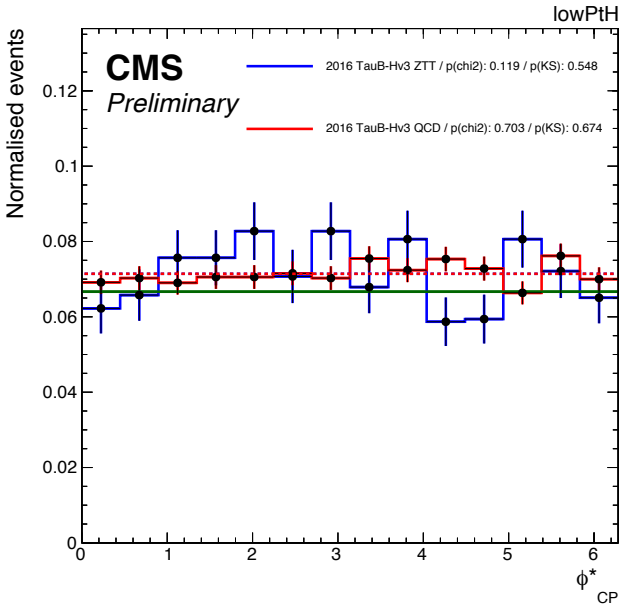
Overall, backgrounds look flat in both data and MC. Can merge bins to gain bin-by-bin statistics.

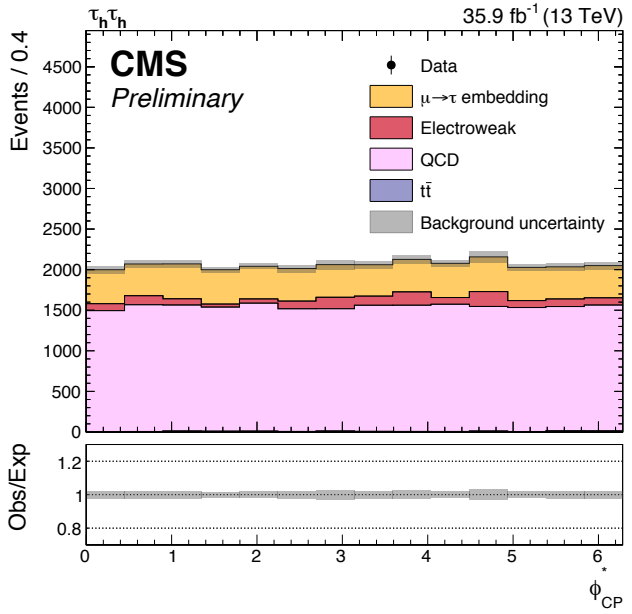
inclusive


inclusive
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.437 / p(KS): 0.853
2016 TauB-Hv3 QCD / p(chi2): 0.721 / p(KS): 0.727
Normalised events
0.12
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
4
5
6
ϕ^*_{CP}

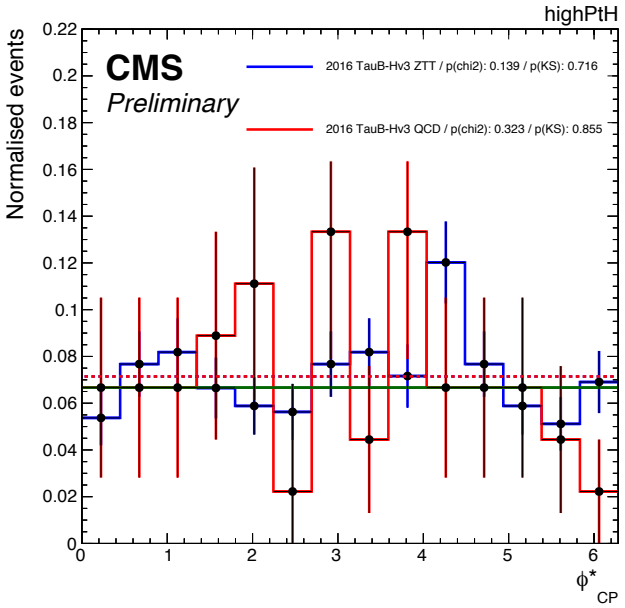

$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
5000
4000
3000
2000
1000
0
Obs/Exp
1.2
1
0.8
0
1
2
3
4
5
6
ϕ^*_{CP}

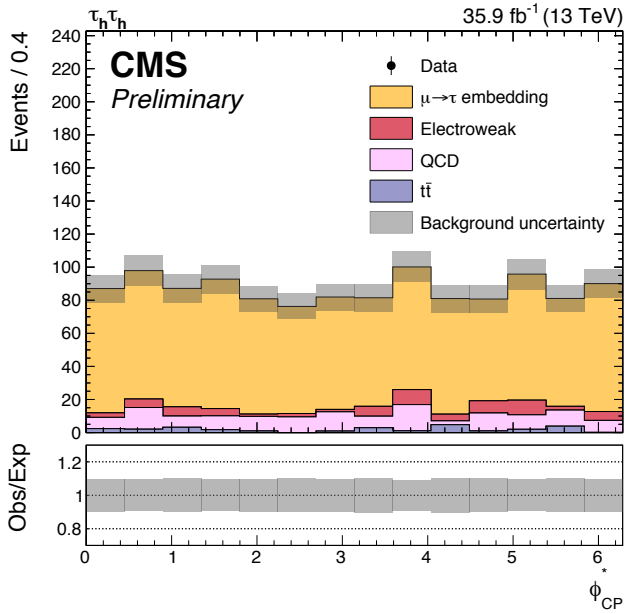
pt_tt < 150



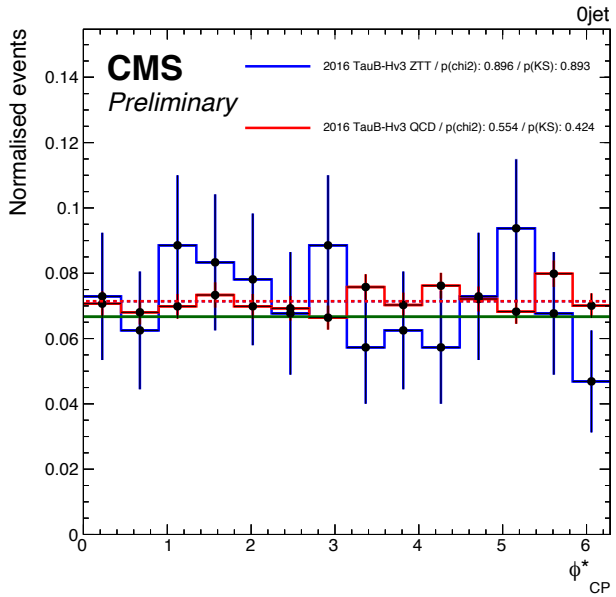


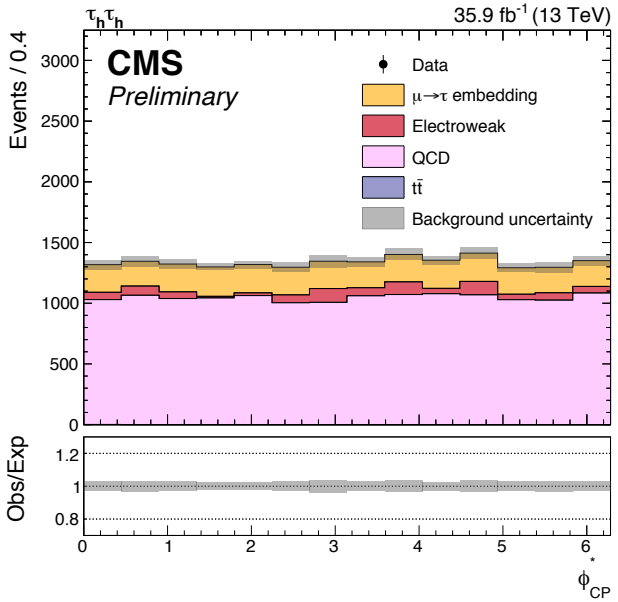
pt_tt > 150



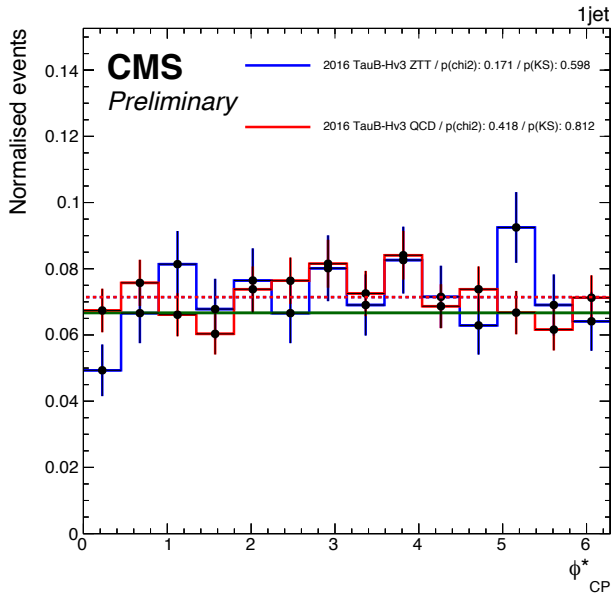


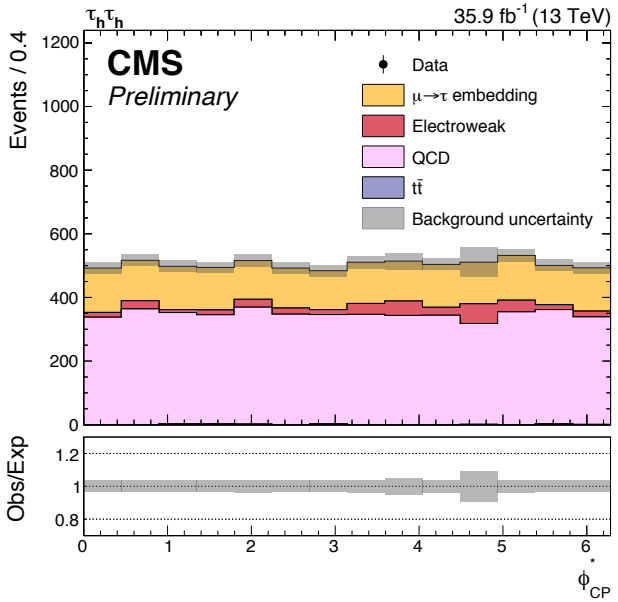
0jet


0jet
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.896 / p(KS): 0.893
2016 TauB-Hv3 QCD / p(chi2): 0.554 / p(KS): 0.424
Normalised events
ϕ^*_{CP}


$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
Obs/Exp
ϕ^*_{CP}

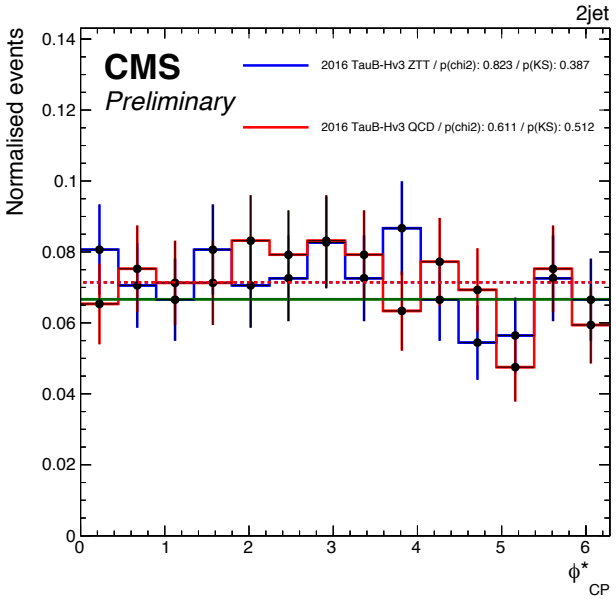
1jet

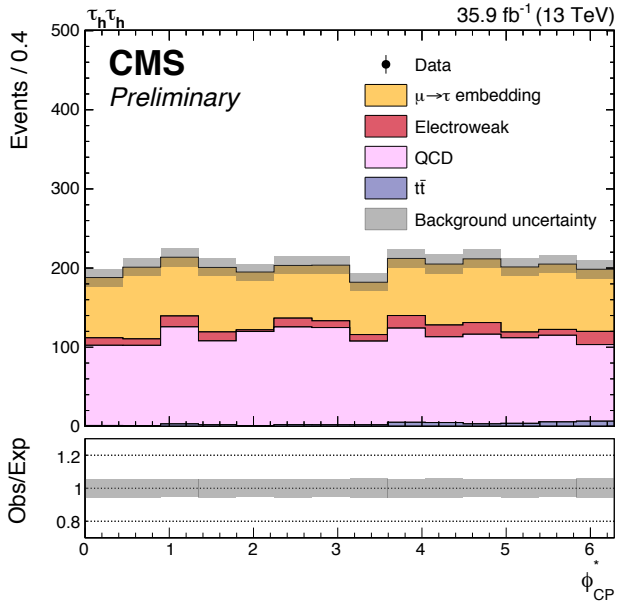

1jet
CMS
Preliminary
2016 TauB-Hv3 ZTT / p(chi2): 0.171 / p(KS): 0.598
2016 TauB-Hv3 QCD / p(chi2): 0.418 / p(KS): 0.812
Normalised events
ϕ^*_{CP}


$\tau_h\tau_h$
35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Data
$\mu\rightarrow\tau$ embedding
Electroweak
QCD
$t\bar{t}$
Background uncertainty
Events / 0.4
Obs/Exp
ϕ^*_{CP}

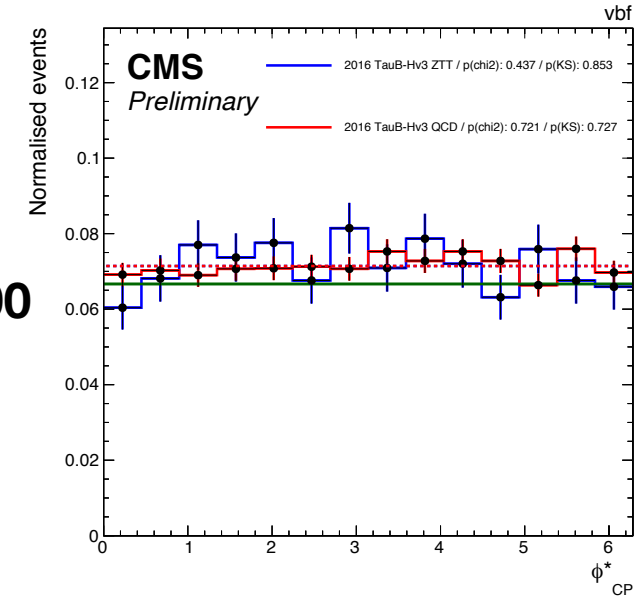
2jet

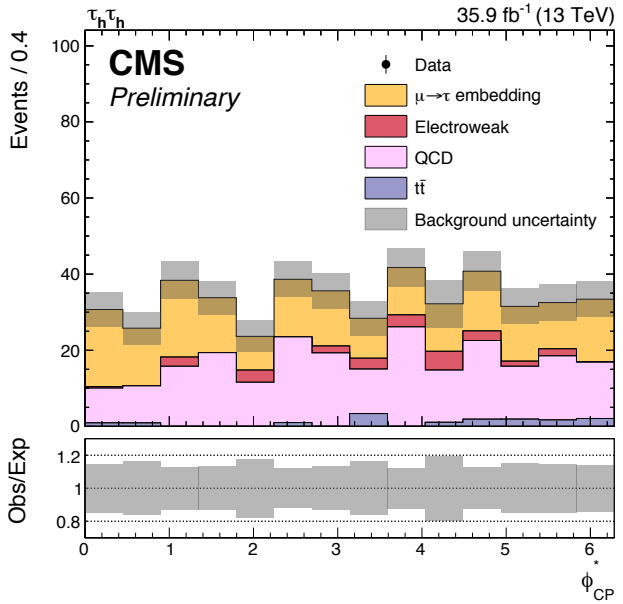
dijet





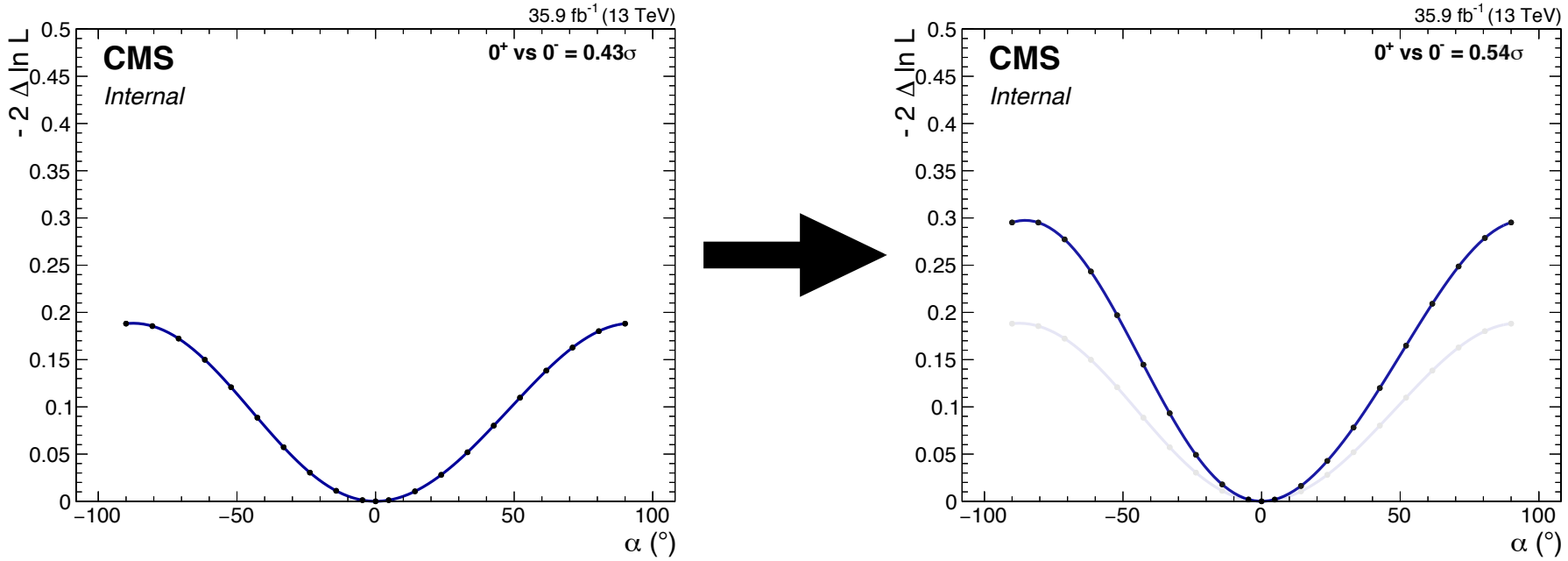
dijet + mjj > 500





Exp. Sensitivity with Decay Mode MVA

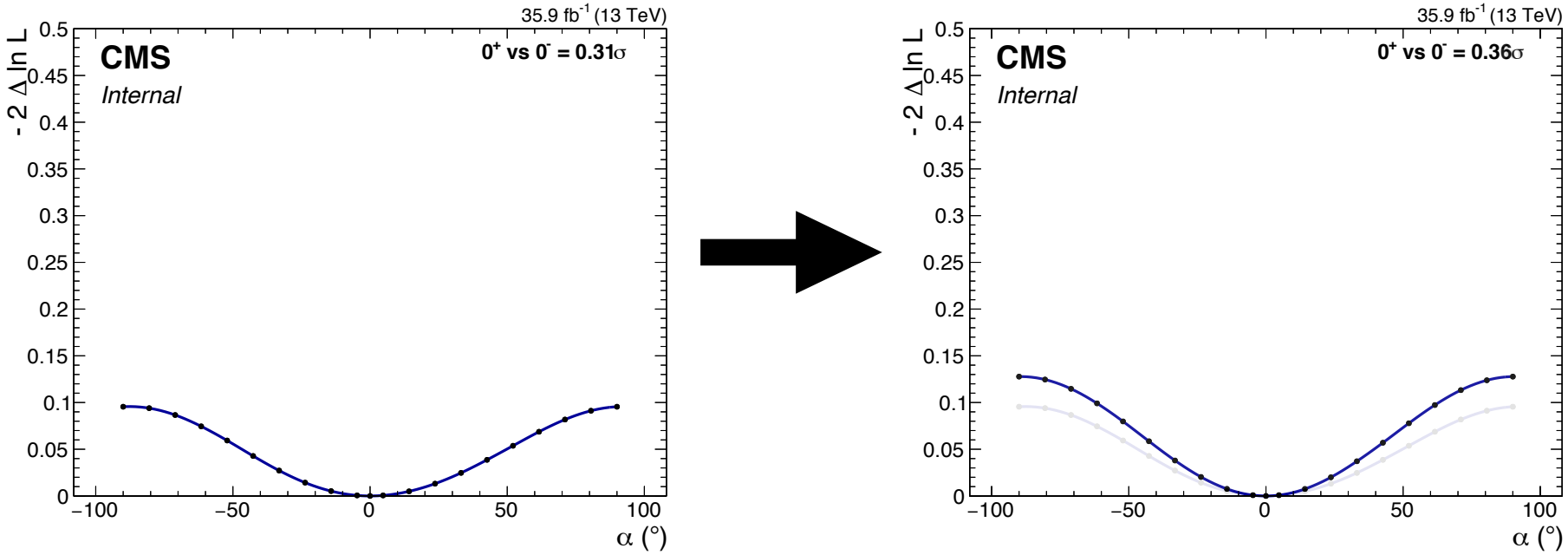
$\rho\rho$ channel sensitivity gains by about 25%.



2016 (80X) only

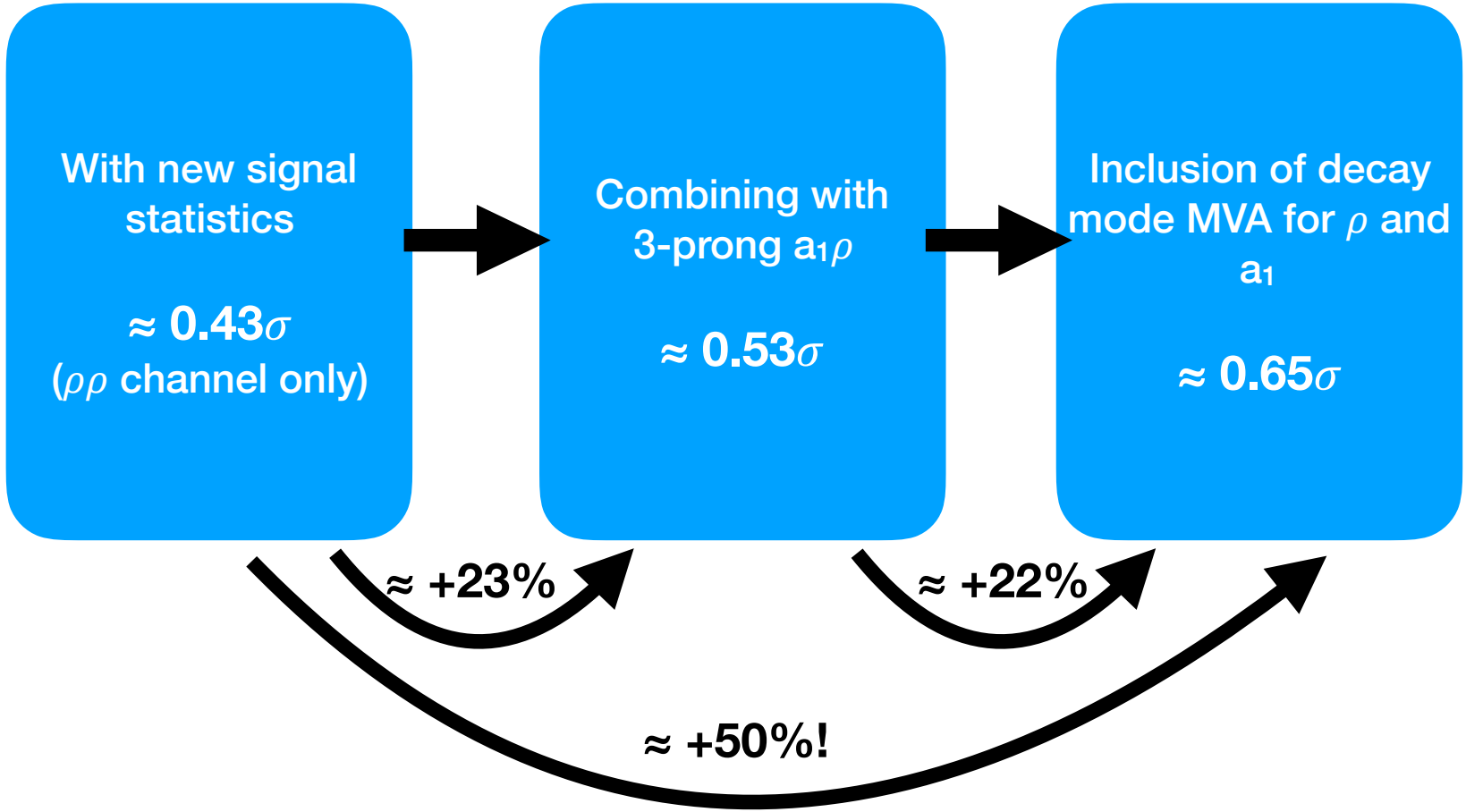
Exp. Sensitivity with Decay Mode MVA

$a_1\rho$ channel sensitivity gains by about 16%.



2016 (80X) only

Summary of Sensitivity Studies



NB. 2016 80X data/MC only, full τ + JES uncertainty model used.

Private 2016 Signal Samples

Inclusive samples with 5M events:

/GluGluHToTauTau_M125_13TeV_powheg_pythia8_nospinner/dwinterb-GluGluHToTauTau_M125_13TeV_powheg_pythia8_nospinner-miniAOD-28028af67189b3de7224b79195bd0e1d/USER

/VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner/dwinterb-VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner-miniAOD-28028af67189b3de7224b79195bd0e1d/USER

With gen filters with 20M events:

/GluGluHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter/dwinterb-GluGluHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter-miniAOD-28028af67189b3de7224b79195bd0e1d/USER

/VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter-v2/dwinterb-VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter-miniAOD-28028af67189b3de7224b79195bd0e1d/USER

/VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter-v2-ext/dwinterb-VBFHToTauTau_M125_13TeV_powheg_pythia8_nospinner-filter-ext-miniAOD-28028af67189b3de7224b79195bd0e1d/USER