$A \to Zh \to ll\tau\tau$

Coffea Users Meeting

Gage DeZoort, Jaana Heikkilae, Alexis Kalogeropoulos,
Dan Marlow, Diwakar Vats, Yiwen Wen

11/22/2021

Theory and Motivation

- Two Higgs-doublet models (2HDMs) extend the SM Higgs sector with an additional Higgs doublet
  - The doublets (ϕ_1, ϕ_2) undergo SSB as usual, acquiring non-zero vevs v_1, v_2

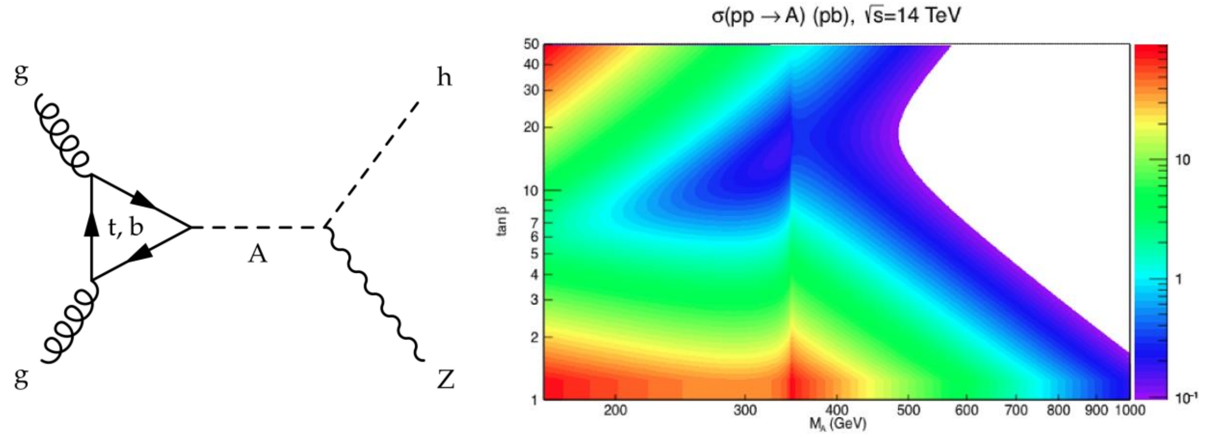
$\tan(\beta) = v_2/v_1$

$(v_1)^2 + (v_2)^2 \approx (246\ GeV)^2$

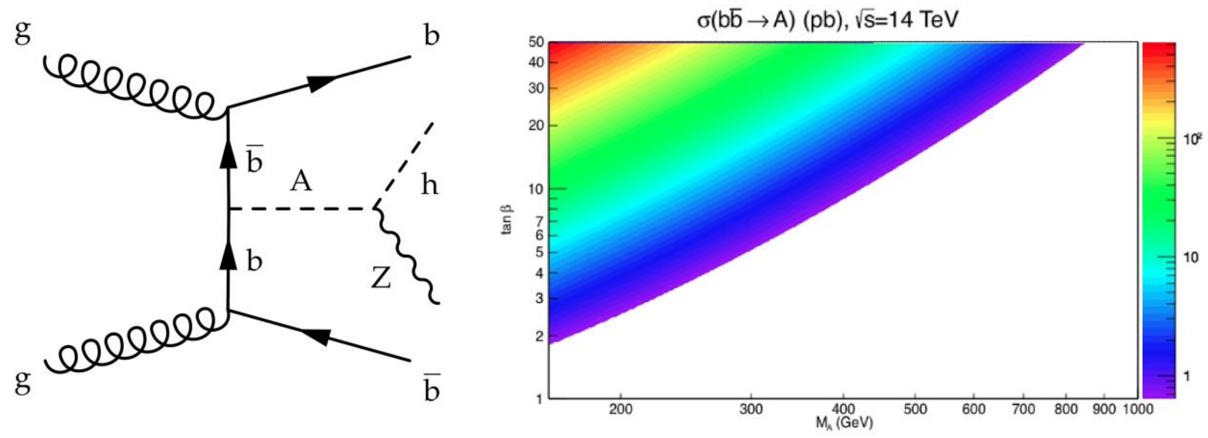
  - The theory yields mass eigenstates corresponding to 5 physical scalar bosons:

H^+, H^-, A, h, H
(2HDM Higgs mass eigenstates)

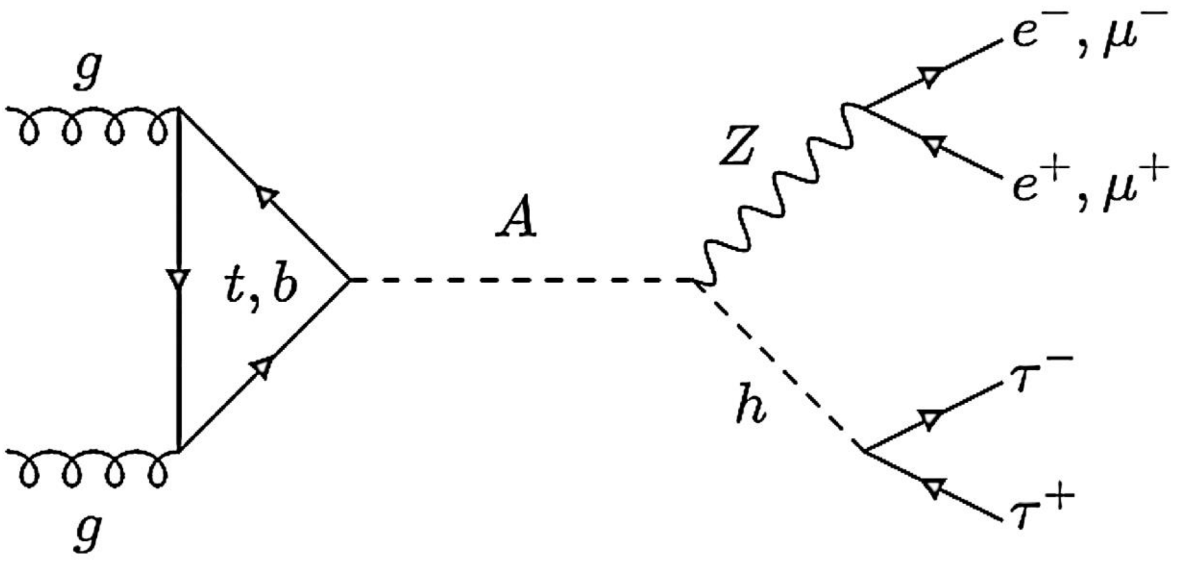
  - The A is a CP-odd (pseudoscalar), neutral Higgs boson
- We consider A masses in the range 220 GeV (threshold for Zh) to 2000 GeV



Gluon-gluon fusion (ggA) dominates for low/moderate tan(β)



Production in association with a bb pair (bbA) becomes relevant for larger tan(β)



Notation:

Final states: $l_1 l_2 \tau_1 \tau_2 \rightarrow \mathrm{leg1},\ \mathrm{leg2},\ \mathrm{leg3},\ \mathrm{leg4}$

$\mathrm{leg\ 1} \rightarrow$ leading light lepton

$\mathrm{leg\ 2} \rightarrow$ sub-leading light lepton

$\mathrm{leg\ 3} \rightarrow$ leading τ_h (in the $\tau_h\tau_h$ final state)

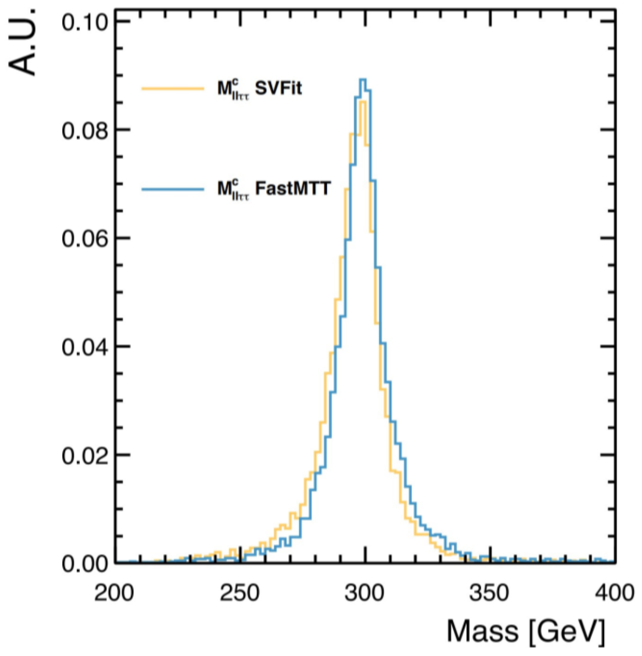
$\mathrm{leg\ 4} \rightarrow$ sub-leading τ_h (in the $\tau_h\tau_h$ final state)

Physics Signatures

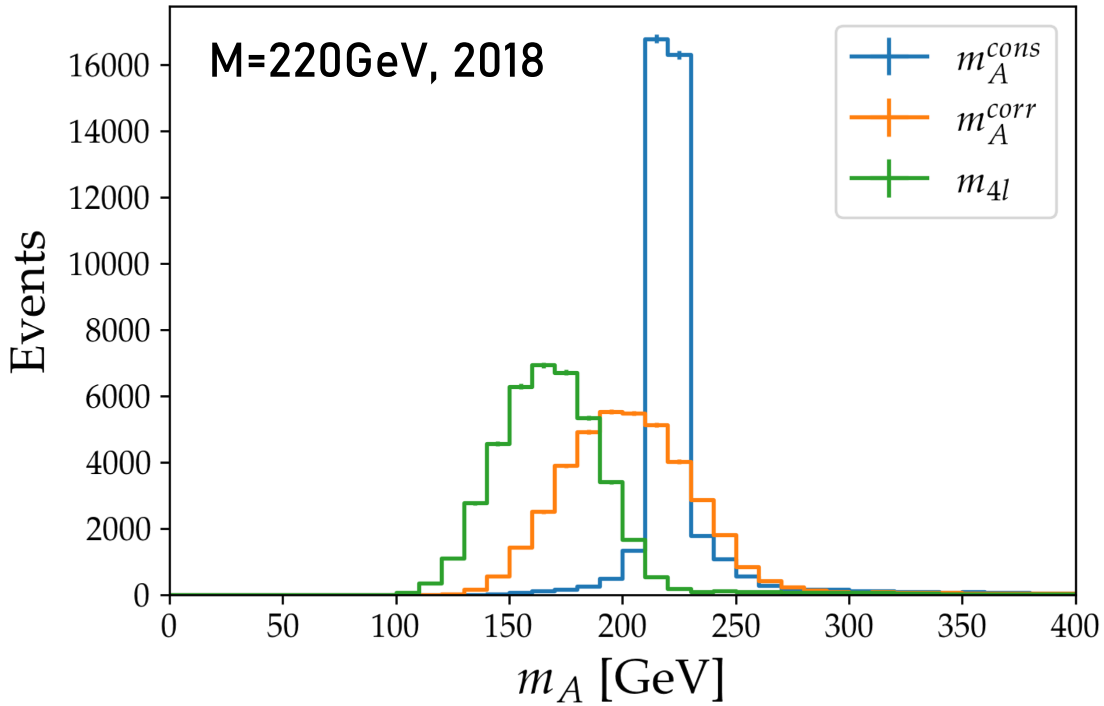
- Here, we consider $4l$ final states corresponding to the decays:
  - $Z \rightarrow ll \quad (l = e, \mu)$
  - $\mathrm{h} \rightarrow \tau\tau \quad (\tau = \tau_e, \tau_\mu, \tau_h)$
- We consider 8 channels:

$ee\tau_h\tau_h$	$\mu\mu\tau_h\tau_h$
$ee\tau_\mu\tau_h$	$\mu\mu\tau_\mu\tau_h$
$ee\tau_e\tau_h$	$\mu\mu\tau_e\tau_h$
$ee\tau_e\tau_\mu$	$\mu\mu\tau_e\tau_\mu$

- Identify events via
  - Single lepton triggers
  - Double Lepton triggers



Comparison between SVFit and FastMTT mass resolution (w/ constraint)



Mass Fits

- Taus decay weakly, and therefore always produce at least 1 neutrino ⇒ **Detector measurements do not fully constrain the kinematics of tau decay**
- We employ the SVfit and FastMTT algorithm to reconstruct the di-tau invariant mass:

corrected $4l$ mass with constraint

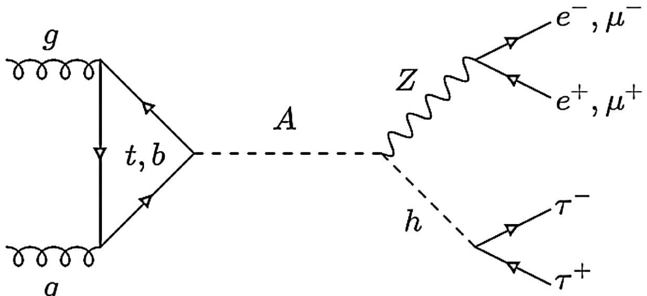
$$(m_A^{cons})^2 = \left(l_1 + l_2 + \tau_1^{fit} + \tau_2^{fit}\right)^2 \text{ s.t. } \left|\left(\tau_1^{fit} + \tau_2^{fit}\right)^2 - 125\right| < 1\ GeV$$

corrected $4l$ mass

$$(m_A^{corr})^2 = \left(l_1 + l_2 + \tau_1^{fit} + \tau_2^{fit}\right)^2$$

raw $4l$ mass

$$(m_{4l})^2 = (l_1 + l_2 + \tau_1 + \tau_2)^2$$



Signal Process

$ee\tau_h\tau_h$	$\mu\mu\tau_h\tau_h$
$ee\tau_\mu\tau_h$	$\mu\mu\tau_\mu\tau_h$
$ee\tau_e\tau_h$	$\mu\mu\tau_e\tau_h$
$ee\tau_e\tau_\mu$	$\mu\mu\tau_e\tau_\mu$

8 Final States

ZZ4l, ttbarV, SM, WWZ, WZZ, ZZZ

Estimate with MC; m_{tt}^{fit} cuts

Irreducible Background

Z+jets WZ+jets ttbar

Estimate via fake factor method; L_T cuts

Reducible Background

Mass of the 4l final state where the ditau mass is constrained within some window of the SM Higgs mass

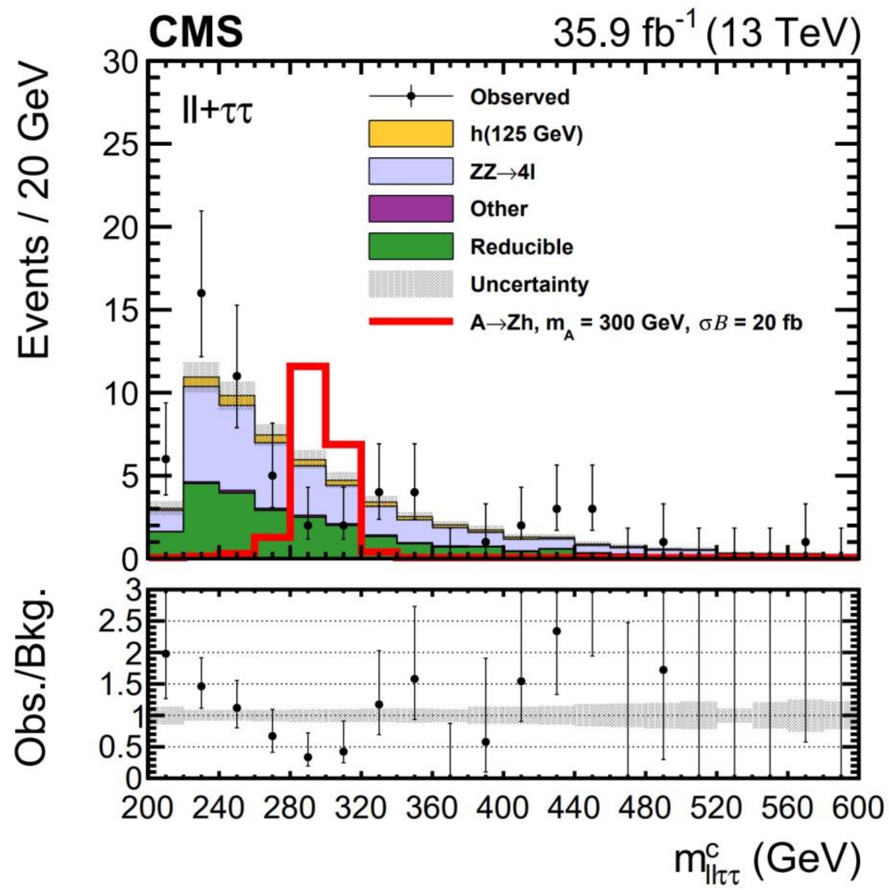
$$(m_A^{cons})^2 = \left(l_1 + l_2 + \tau_1^{fit} + \tau_2^{fit}\right)^2$$

$$\text{s.t. } \left|\left(\tau_1^{fit} + \tau_2^{fit}\right)^2 - 125\right| < \delta\ [GeV]$$

Discriminating Variable

Analysis Strategy

- HIG-18-023 signal vs. background plot:



Organization

- Previous Analysis:
  "Search for a heavy pseudoscalar Higgs boson …" (arXiv:1910.11634)
- AZh Run 2 TWiki:
  https://twiki.cern.ch/twiki/bin/viewauth/CMS/AZh-Htautau-Run2

PRINCETON UNIVERSITY

- Personnel:
  - Dan Marlow (Professor)
  - Alexis Kalogeropoulos (Postdoc)
  - Gage DeZoort (PhD Student)
- Analysis based on NanoAOD (nAOD) data
- Coffea-based preselection framework

Princeton Framework

DESY. NISER

- DESY Personnel:
  - Yiwen Wen (Postdoc)
- NISER Personnel:
  - Diwakar Vats (PhD Student)
- Analysis based on MiniAOD (mAOD) data
- CMSSW-based preselection framework

DESY Framework

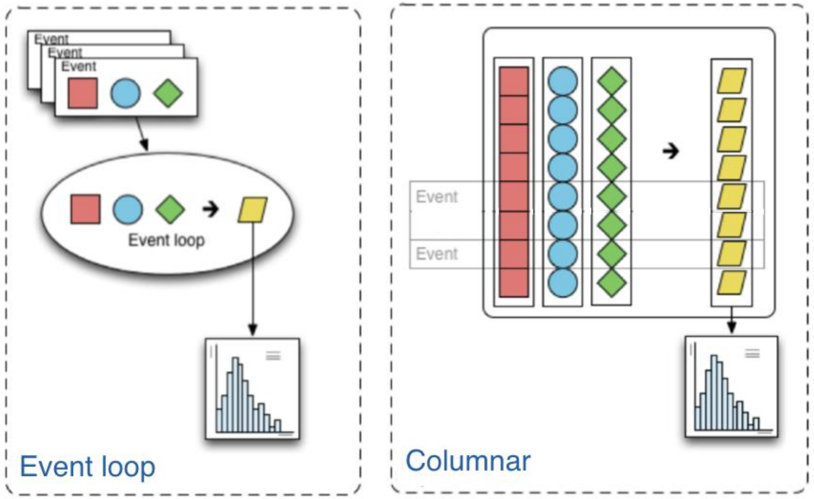
University of Zurich UZH

- Personnel:
  - Jaana Heikkilae (Postdoc)
- Expertise from the 2016 analysis

Coffea Framework

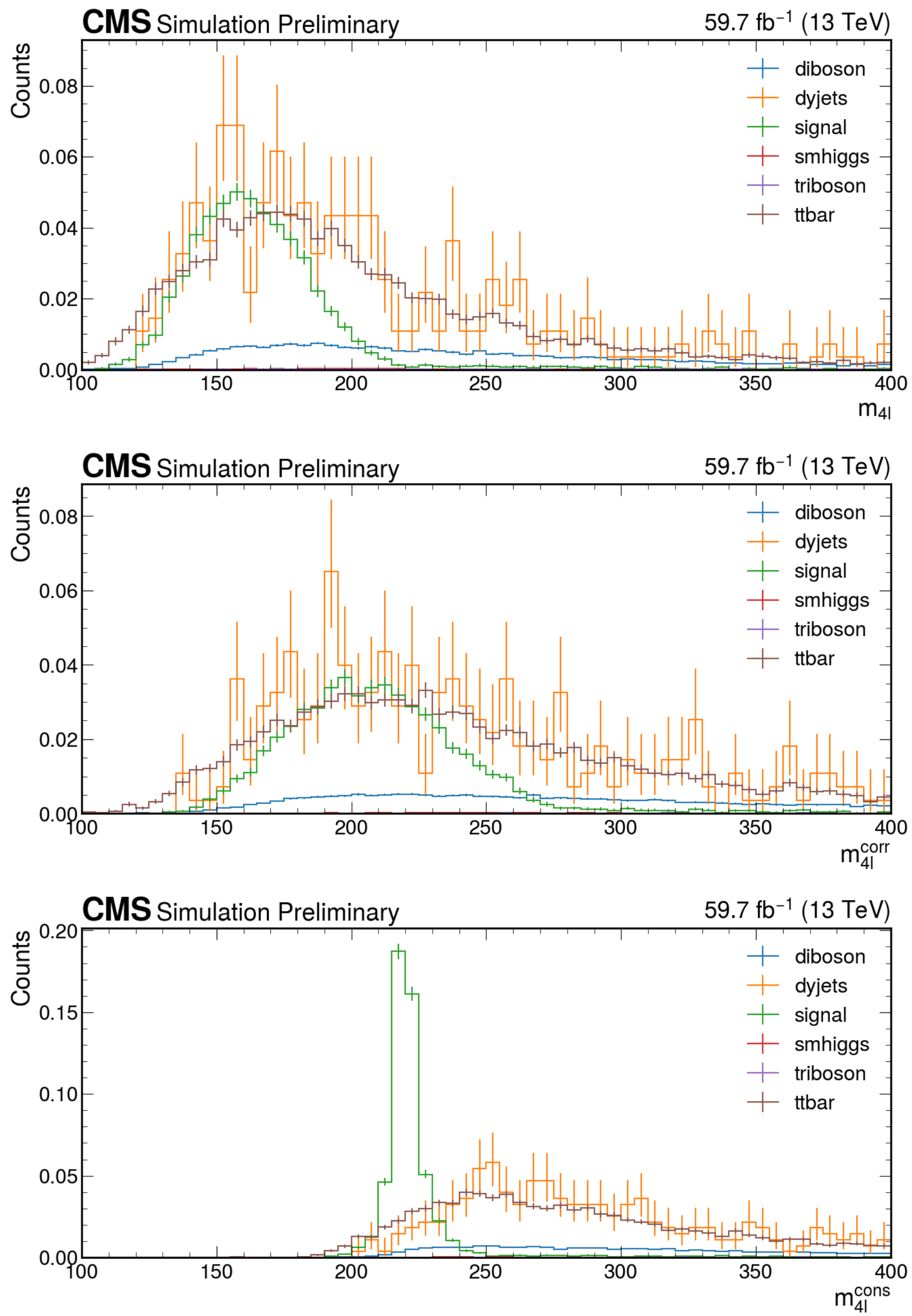
- Personnel:
  - Dan Marlow (Professor)
  - Alexis Kalogeropoulos (Postdoc)
  - Gage DeZoort (PhD Student)
- Analysis based on NanoAOD (nAOD) data
- Coffea-based preselection framework https://github.com/GageDeZoort/AZh_Princeton

Princeton Framework



Coffea Framework

- Run via lpcjobqueue:
  - 2018 MC only
  - Signal sample: A mass 220 GeV
  - 560,332,448 events
  - Typically takes 30-60mins



FastMTT in Python

- Algorithm:
  - Key assumptions: $m_\tau \approx 0$, collinear approximation (tau decay products decay in-line with original tau momentum) used to derive: $m_{\tau\tau} = m_{vis}/\sqrt{(X_1 X_2)}$
  - Parametrize each tau decay as a visible system and an invisible system with the following parameters:
    - Visible: $E^1_{vis}, \vec{p}^{\,1}_{vis}, E^2_{vis}, \vec{p}^{\,2}_{vis}$
    - Invisible: $X_1, X_2, \phi_1, \phi_2, m^1_{\nu\nu}, m^2_{\nu\nu}$
      - $X_i \rightarrow E^i_{vis}/E^i_\tau$
      - $\phi_i \rightarrow$ azimuth angle of tau direction around visible momenta
- Implementation in Python, compared to C++ implementation
- Tested on 5000 AZh events at mass 220 GeV
  - Python Implementation: 7ms / event
  - C++ Implementation: 1.5ms / event

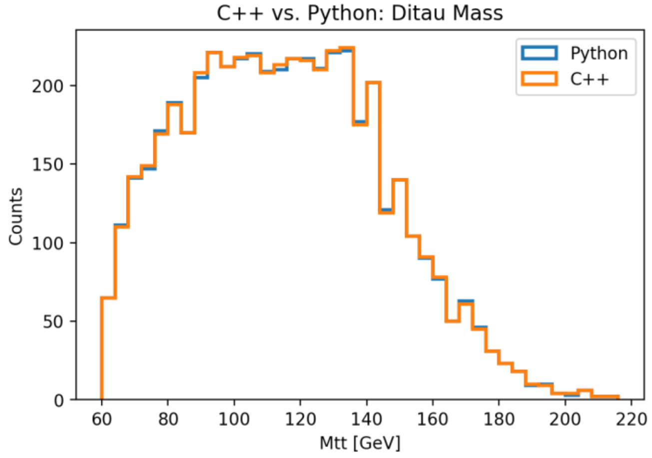
```python
# initialize Higgs--> tau + tau decays, tau decay types
N = len(pt_1)
m_tt_vis = np.zeros(N, dtype=np.float32)
m_tt_opt = np.zeros(N, dtype=np.float32)
for i in range(N):
    if (decay_type_1[i] != 2): m1 = mass_dict[decay_type_1[i]]
    else: m1 = mass1[i]
    if (decay_type_2[i] != 2): m2 = mass_dict[decay_type_2[i]]
    else: m2 = mass2[i]
    leg1 = vec.obj(pt=pt_1[i], eta=eta_1[i], phi=phi_1[i], mass=m1)
    leg2 = vec.obj(pt=pt_2[i], eta=eta_2[i], phi=phi_2[i], mass=m2)
    p4_vis = leg1 + leg2 # visible ditau 4-vector
    m_vis = p4_vis.mass # visible ditau mass
    m_tt_vis[i] = m_vis
    m_vis_1 = leg1.mass
    m_vis_2 = leg2.mass

    if (decay_type_1[i] == 2 and m_vis_1 > 1.5): m_vis_1 = 0.3
    if (decay_type_2[i] == 2 and m_vis_2 > 1.5): m_vis_2 = 0.3

    metcovinv_xx, metcovinv_yy = metcov_yy[i], metcov_xx[i]
    metcovinv_xy, metcovinv_yx = -metcov_xy[i], -metcov_yx[i]
    metcovinv_det = (metcovinv_xx*metcovinv_yy -
                     metcovinv_yx*metcovinv_xy)

    if (metcovinv_det<1e-10):
            print("Warning! Ill-conditioned MET covariance at event inde
            continue

    met_const = 1/(2*math.pi*math.sqrt(metcovinv_det))
    min_likelihood, x1_opt, x2_opt = 999, 0, 0
    mass_likelihood, met_transfer = 0, 0
    for x1 in np.arange(0, 1, 0.01):
        for x2 in np.arange(0, 1, 0.01):
            x1_min = min(1, math.pow((m_vis_1/m_tau), 2))
            x2_min = min(1, math.pow((m_vis_2/m_tau),2))
            if ((x1 < x1_min) or (x2 < x2_min)):
```

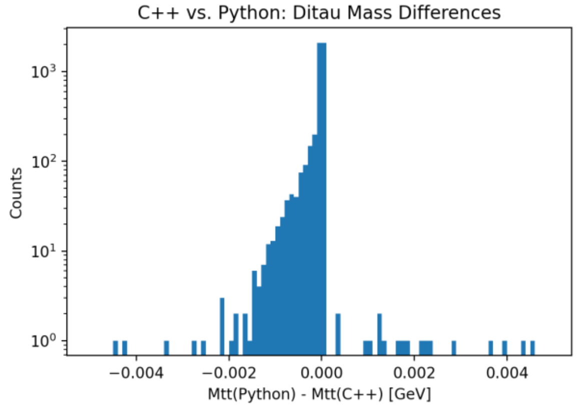
loop over all ditau candidates

tau 4-momentum vectors represented by VECTOR

brute-force likelihood scan over x_1, x_2

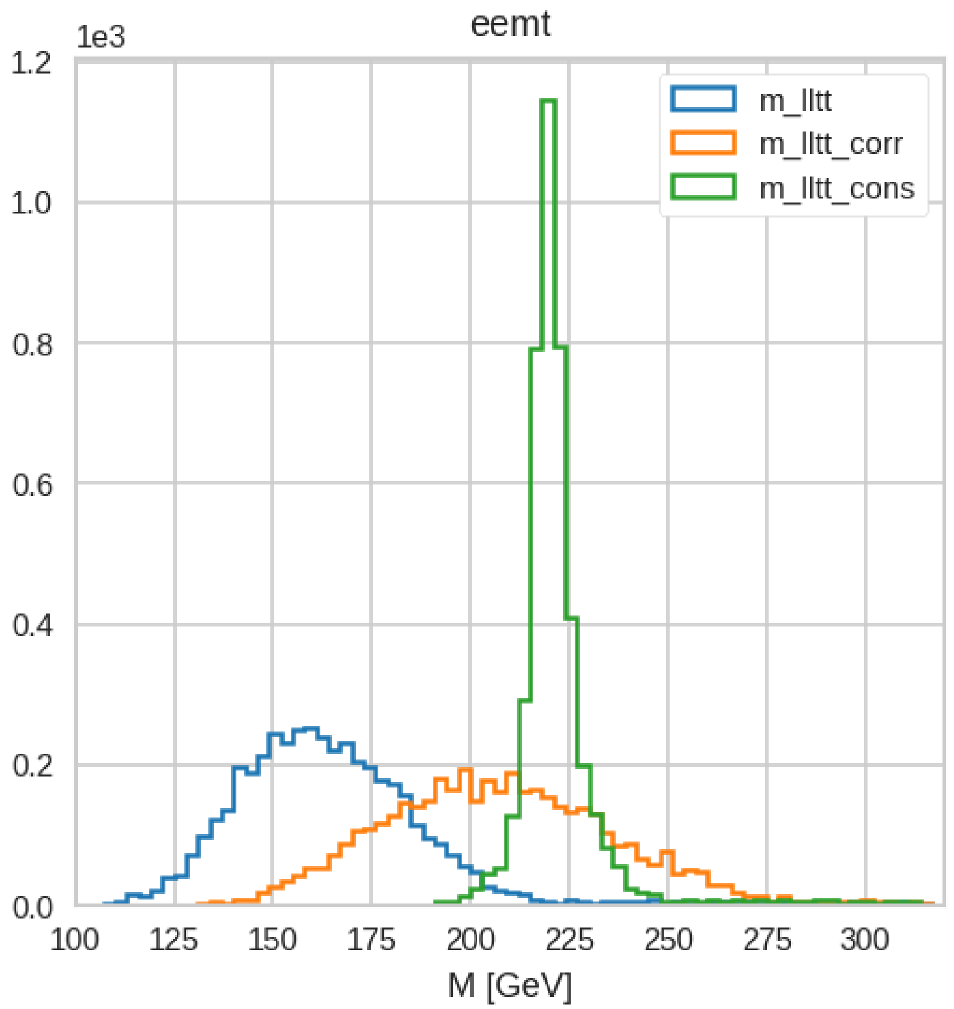
accelerated via Numba's just-in-time (JIT) compilation





FastMTT in Python

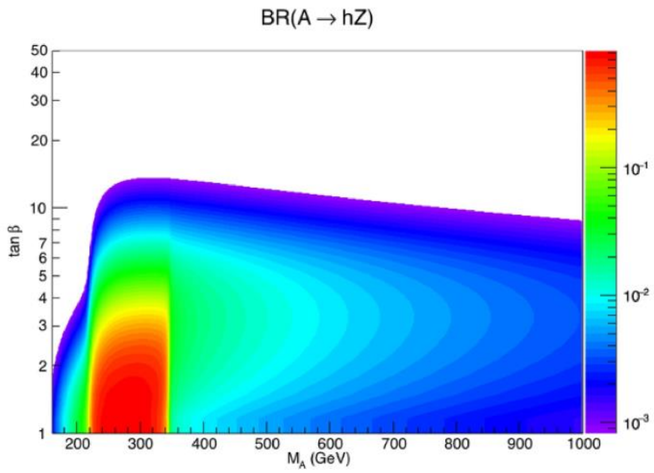
- Implemented to run at pre-selection stage of Coffea framework with and without mass constraint
- Plot: $ee\mu\tau$ final state with sync selections run on a 2018 ggA signal sample at $m_A = 220$ GeV
  - m_{lltt} is the raw 4-lepton mass
  - m_{lltt}^{corr} is the corrected 4-lepton mass
  - m_{lltt}^{cons} is the constrained 4-lepton mass (ditau mass constrained to 125 ± 1 GeV)



Thank you!

Previous Constraints

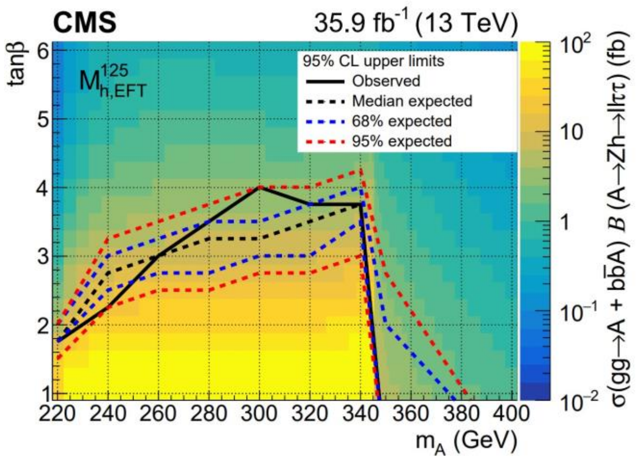
- The Minimal Supersymmetric SM (MSSM) is a type II 2HDM; key scenarios:
  - hMSSM: require m_h = 125 GeV s.t. Higgs sector is parametrized by only $\tan(\beta)$ and m_A
  - $M_{h,EFT}^{125}$: derived from 2HDM EFT, SUSY mass scale makes m_h = 125 GeV accessible at low $\tan(\beta)$



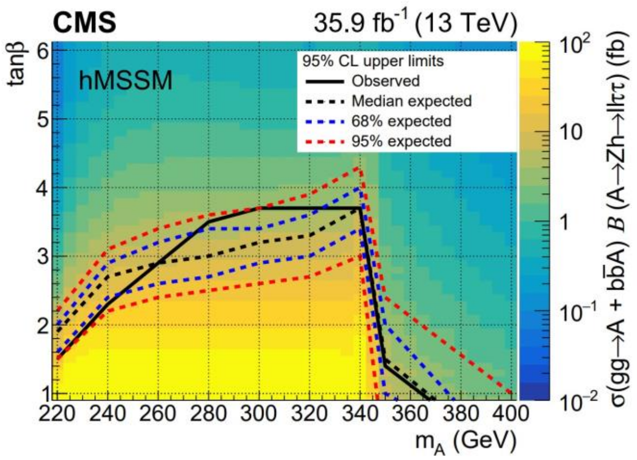
In the hMSSM, $\mathrm{m_h} = 125$ GeV is fixed and A decays to Zh dominate for low $\tan(\beta)$

(see doi:10.1007/JHEP06(2015)168)

Model-dependent limits constrain low $\tan(\beta)$ at m_A = 220-350 GeV

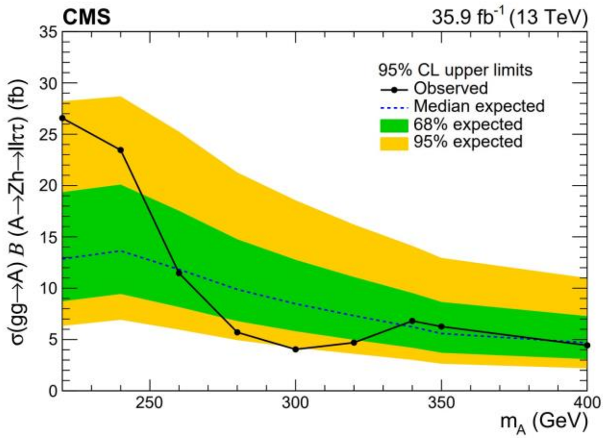


$M_{h,EFT}^{125}$ model-dependent limits set in HIG-18-023



hMSSM model-dependent limits set in HIG-18-023

Model-independent limits (all channels combined) are consistent with the absence of signal



(see doi:10.1007/JHEP03(2020)065)

di-Tau Mass Fits

- Taus decay weakly, and therefore always produce at least 1 neutrino
  $\Rightarrow$ **Detector measurements do not fully constrain the kinematics of tau decay**
- We employ the SVfit algorithm to reconstruct the di-tau invariant mass:

  https://github.com/SVfit/ClassicSVfit

- ClassicSVfit delivers:
  - Reconstructed tau p_T, η, ϕ, as well as $m_{\tau\tau}$ and $m_{T\tau\tau}$
  - Capability to perform a constrained mass fit, where $m_{\tau\tau} = 125$ GeV is fixed

  https://arxiv.org/abs/1603.05910

L. Bianchini, B. Calpas, J. Conway, A. Fowlie, L. Marzola, L. Perrini, C. Veelken, "Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of tau leptons using matrix element techniques", Nucl. Instrum. Meth. A 862 (2017) 54

Initial 4-vectors:

$$l_1^\mu = (p_{T,1}, \eta_1, \phi_1, m_1)$$
$$l_2^\mu = (p_{T,2}, \eta_2, \phi_2, m_2)$$
$$\tau_1^\mu = (p_{T,3}, \eta_3, \phi_3, m_3)$$
$$\tau_2^\mu = (p_{T,4}, \eta_4, \phi_4, m_4)$$

Initial MET Quantities:

$$E_{T,x}^{miss} = \mathrm{E_T^{miss}} \cos(\phi^{miss})$$
$$E_{T,y}^{miss} = \mathrm{E_T^{miss}} \sin(\phi^{miss})$$
$$MET_{cov} = \begin{pmatrix} MET_{cov}^{00} & MET_{cov}^{01} \\ MET_{cov}^{10} & MET_{cov}^{11} \end{pmatrix}$$

calculate raw $4l$ invariant mass:

$$(m_{ll\tau\tau})^2 = (l_1^\mu + l_2^\mu + \tau_1^\mu + \tau_2^\mu)^2$$

SVFit (unconstrained)

SVFit (constrained)

reconstructed di-tau mass and 4-vectors

$$m_{\tau\tau}^{fit}, (\tau_1^\mu)^{fit}, (\tau_2^\mu)^{fit}$$

$$m_{\tau\tau}^{fit} = 125\ GeV, (\tau_1^\mu)^{fit}, (\tau_2^\mu)^{fit}$$

fitted $4l$ mass

$$(m_{ll\tau\tau}^{fit})^2 = \left(l_1^\mu + l_2^\mu + (\tau_1^\mu)^{fit}, (\tau_2^\mu)^{fit}\right)^2$$

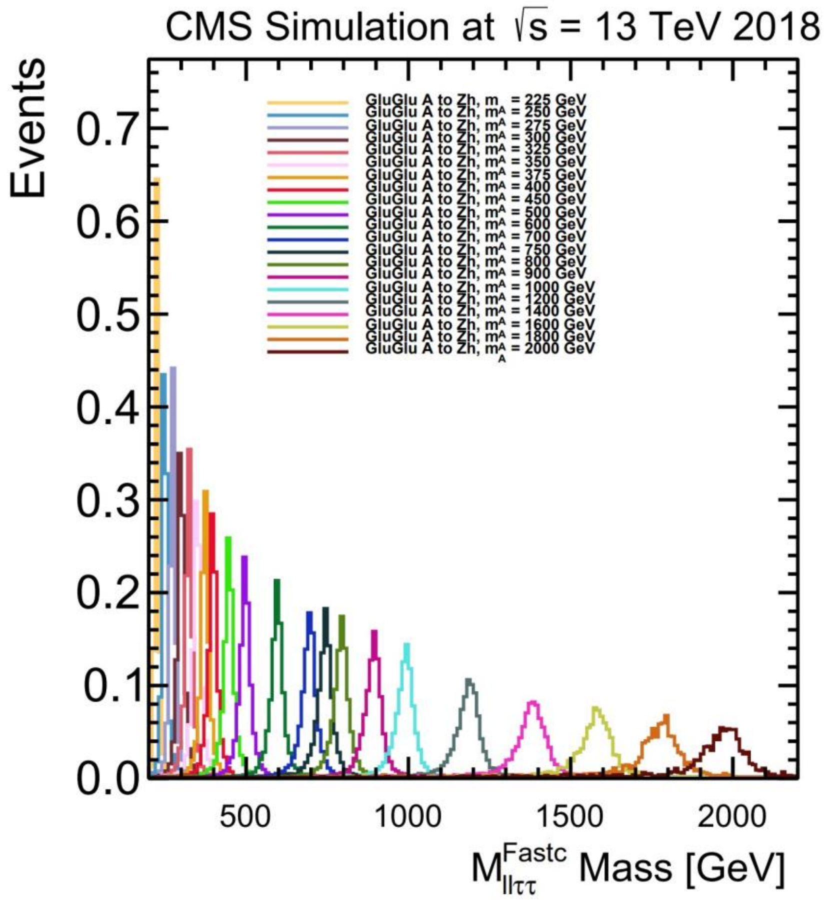
fitted $4l$ mass with constraint

$$(m_{ll\tau\tau}^{constrained})^2 = \left(l_1^\mu + l_2^\mu + (\tau_1^\mu)^{fit}, (\tau_2^\mu)^{fit}\right)^2$$

Status Report

- Completed items:
  - Secured Loop-induced ggA and bbA Run 2 Legacy samples
  - Lepton SFs, trigger SFs, additional corrections
  - SVFit and FastMTT implemented with mass constraints
- In progress:
  - Synchronization in all channels
  - Fake rate calculations
  - Statistical analysis framework
  - Coffea-based framework scale out

GluGlu A mass distributions



N(Princeton)	40801
N(DESY)	40621
N(Princeton only)	237
N(DESY only)	57

Latest sync results in all channels combined

Sync Overview

- Currently synchronizing on a 2018 signal sample (220 GeV) in all 8 channels
- Pileup suppression via PV quality selections
- HLT Trigger Paths

  $Z \rightarrow ee$: ELE35WPTight_Gsf

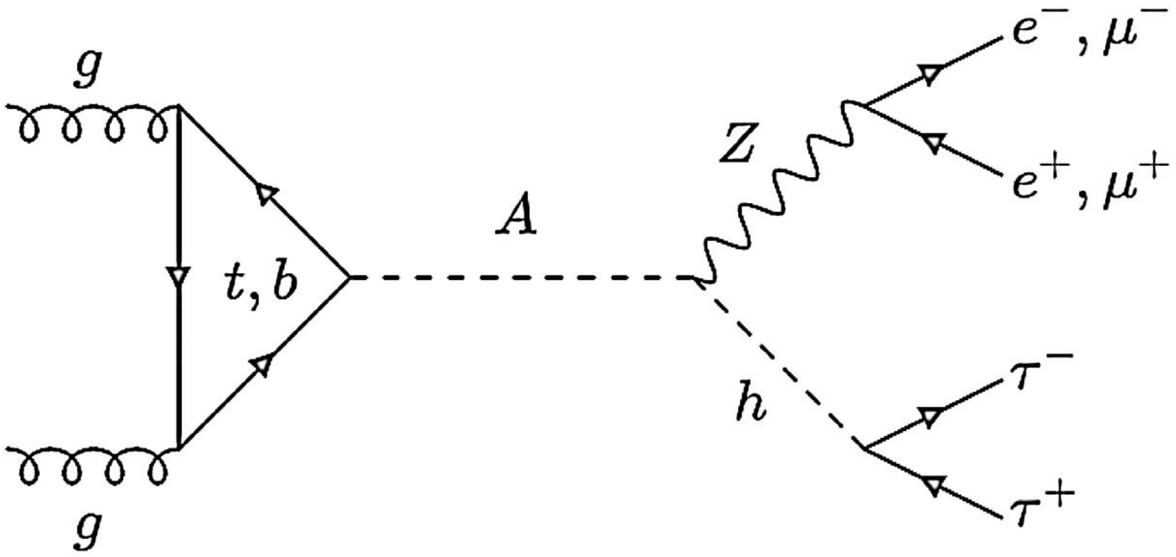
  $Z \rightarrow \mu\mu$: IsoMu27
- Events with any extra leptons are vetoed (e.g. $ee\mu\tau$ must contain exactly 2 loose electrons and 1 loose muon)
- 4l final states are built from loose leptons:
  - τ_h must satisfy dR>0.5 from other final state objects
  - e, μ must satisfy dR>0.3 from other final state objects

Loose Object Selections (nAOD)

Loose e Selections
$\lvert d_{xy}\rvert < 0.045$cm
$\lvert d_z\rvert < 0.2$cm
$p_T > 10$ GeV
$\lvert\eta\rvert < 2.5$
lostHits < 2
convVeto
mvaFall17V2noIso_WP90

Loose μ Selections
$\lvert d_{xy}\rvert < 0.045$ cm
$\lvert d_z\rvert < 0.2$ cm
$p_T > 10$ GeV
$\lvert\eta\rvert < 2.4$
isGlobal or isTracker
looseId or mediumId or tightId

Loose τ_h Selections
$\lvert d_z\rvert < 0.2$cm
$p_T > 20$ GeV
$\lvert\eta\rvert < 2.3$
idDecayModeNewDMs==1
decayMode != 5 and decayMode != 6
idDeepTau2017v2p1VSJet > 0 (VVVLoose)
idDeepTau2017v2p1VSmu > 0 (Loose)
idDeepTau2017v2p1VSe > 3 (VLoose)
lostHits < 2
convVeto
mvaFall17V2noIso_WP90



Z Candidate Selection
(1) Pair OS loose leptons ($e^+e^-, \mu^+\mu^-$)
(2) Select pairs satisfying $60 < m_{ll} < 120$ GeV
(3) Select pair minimizing $\lvert m_{ll} - 91.2 \text{ GeV}\rvert$

di-Tau Candidate Selection
(1) Pair visible τ decay products ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$)
(2) Impose tight DeepTau selections if $e\tau_h$ or $\mu\tau_h$
(3) Select pair maximizing $L_T = (p_T^{(\tau_1)} + p_T^{(\tau_2)})$

Sync Overview

- Z candidates are constructed from loose lepton pairs
- Trigger filter is applied to Z candidate leptons:
  - dR<0.5 from trigger object
  - Trigger object must match corresponding filter bit
  - Lepton p_T > trigger threshold + 1 GeV
- di-Tau candidates are constructed from loose taus and leftover loose leptons

Sync Results

- Much lower disagreement rates
  - 237/40801 ~ 0.5% in Princeton's sample
  - 57/40621 ~ 0.1% in DESY's sample

All Categories	
N(Princeton)	40801
N(DESY)	40621
N(Princeton only)	237
N(DESY only)	57

eeem	
N(Princeton)	1219
N(DESY)	1283
N(Princeton only)	14
N(DESY only)	6

eeet	
N(Princeton)	2910
N(DESY)	2882
N(Princeton only)	38
N(DESY only)	10

eemt	
N(Princeton)	4504
N(DESY)	4504
N(Princeton only)	14
N(DESY only)	14

eett	
N(Princeton)	6516
N(DESY)	6504
N(Princeton only)	23
N(DESY only)	11

mmem	
N(Princeton)	2123
N(DESY)	2103
N(Princeton only)	22
N(DESY only)	2

mmet	
N(Princeton)	4768
N(DESY)	4765
N(Princeton only)	12
N(DESY only)	9

mmmt	
N(Princeton)	7848
N(DESY)	7762
N(Princeton only)	88
N(DESY only)	2

mmtt	
N(Princeton)	10917
N(DESY)	10818
N(Princeton only)	26
N(DESY only)	3