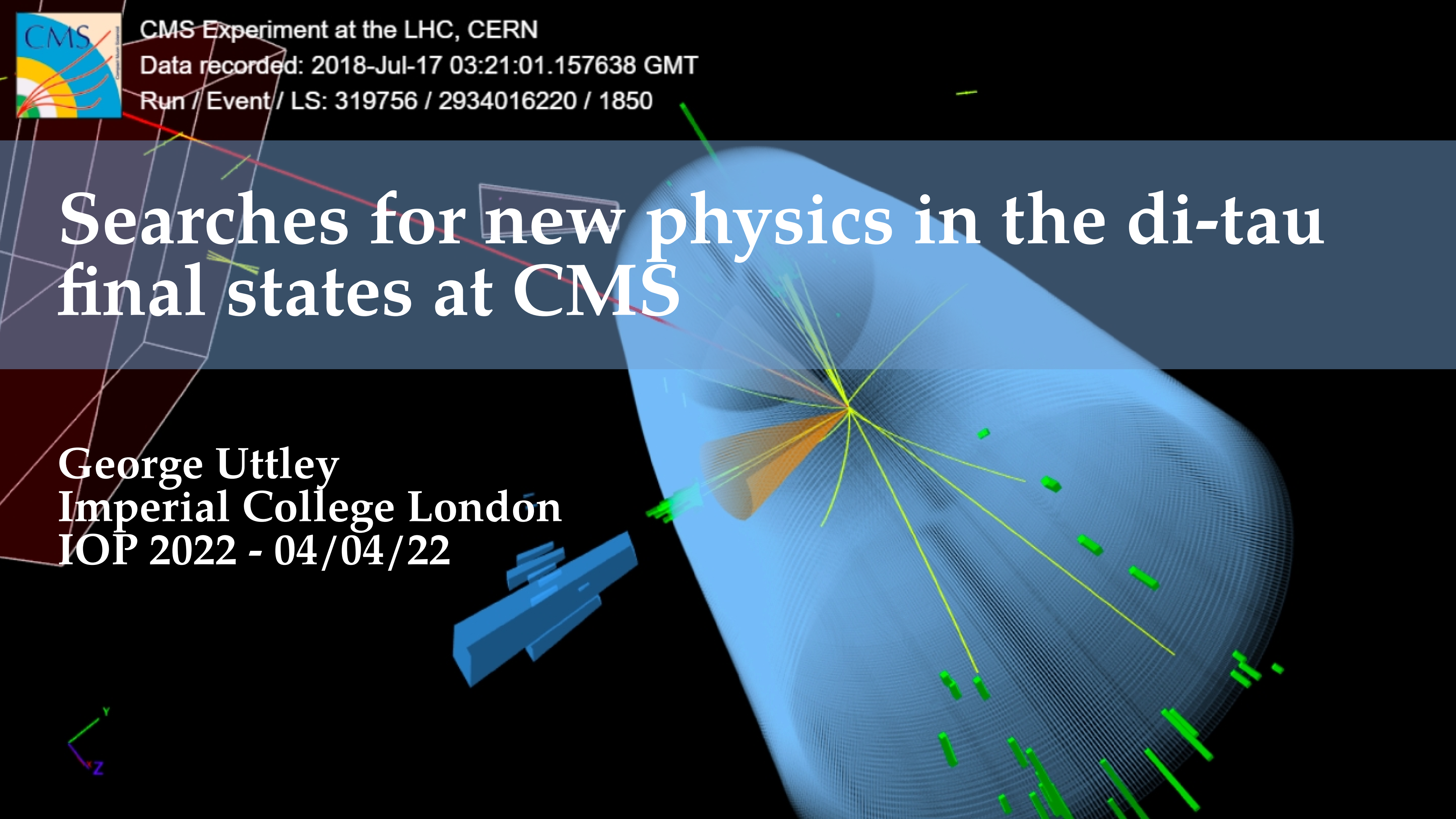


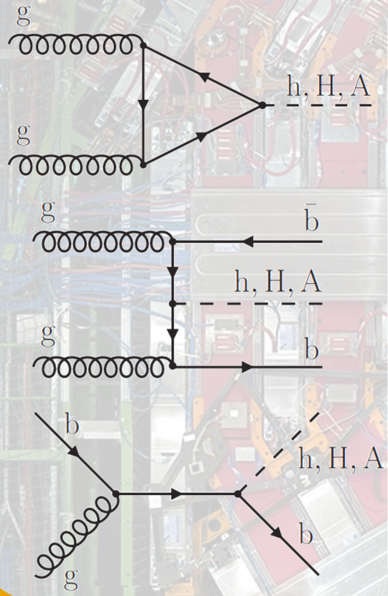
Searches for new physics in the di-tau final states at CMS

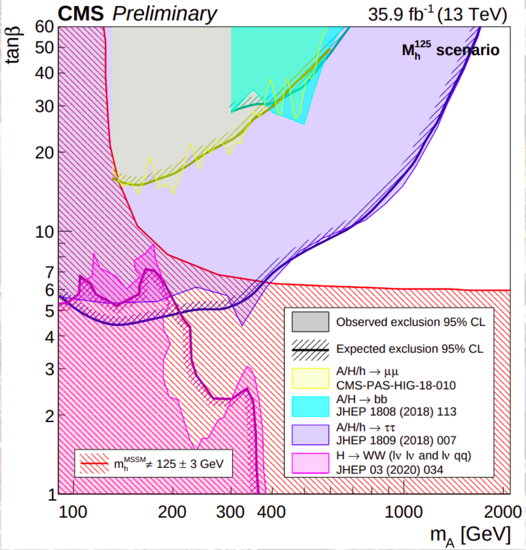
George Uttley
Imperial College London
IOP 2022 - 04/04/22

Introduction

Additional neutral Higgs bosons

Predicted by two Higgs doublet models (including the MSSM). MSSM predicts h, H and A.

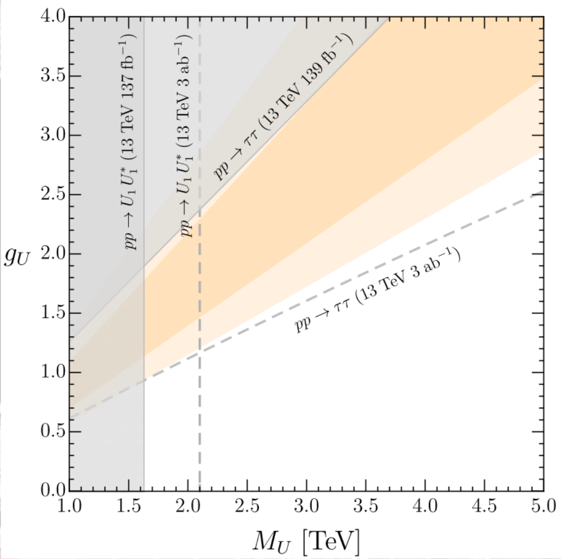




https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryResultsHIG

Vector Leptoquarks

Theorised to simultaneously explain the b anomalies.

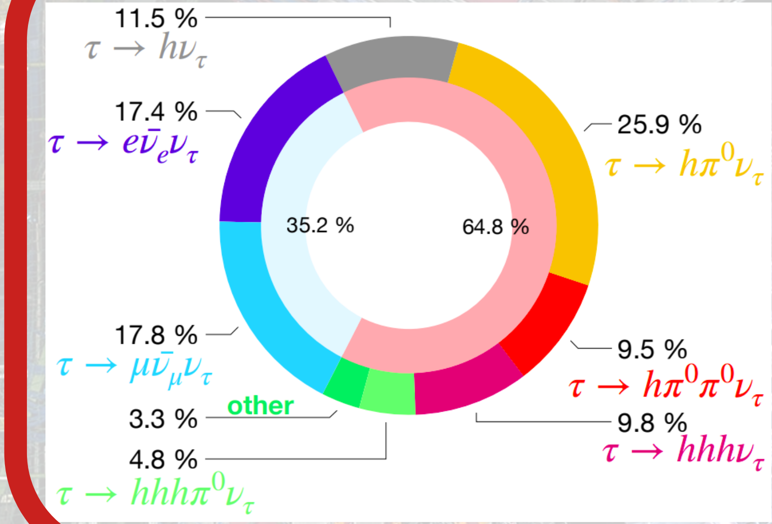


https://arxiv.org/pdf/2103.16558.pdf

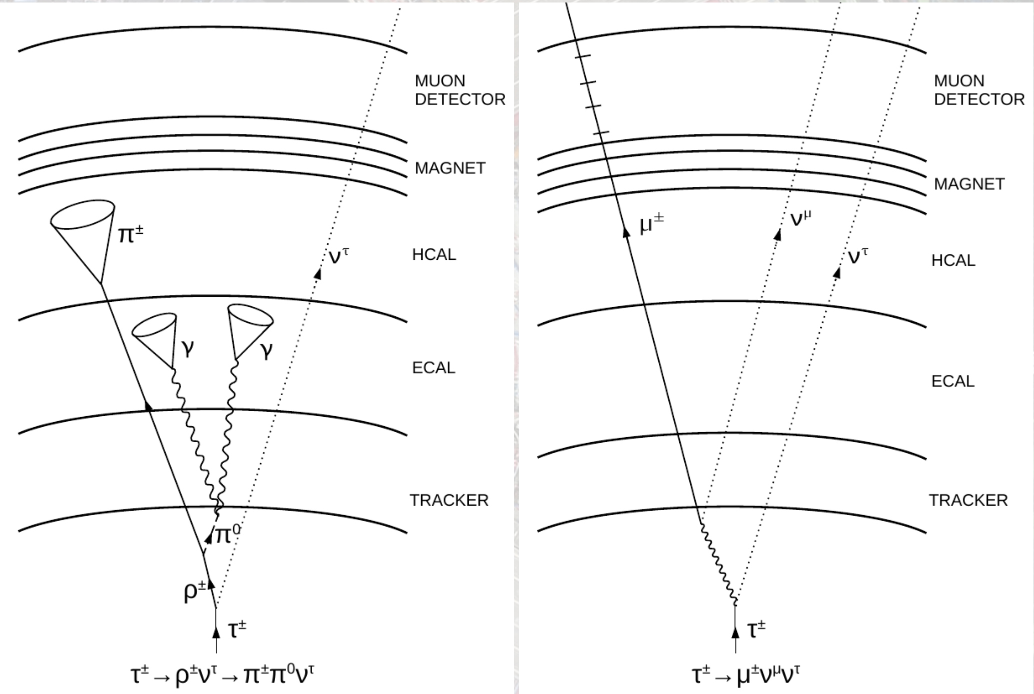
Both theories offer signals in high p_T di-tau tails with possible additional b-jets.

Di-tau Reconstruction

Tau Decay



Finding Taus in CMS Detector



- Taus decay leptonically (**e**,**μ**) or hadronically (**τ_h**).
- We search for 4 di-tau final states: **$\tau_h\tau_h$, $e\tau_h$, $\mu\tau_h$ and eμ**.
- Hadronic taus selected by the DeepTau algorithm.

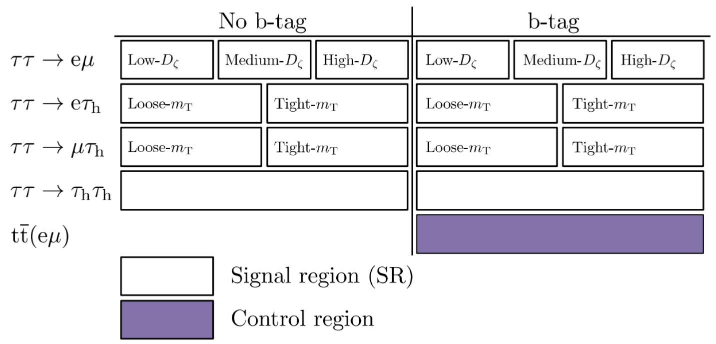
https://arxiv.org/pdf/2201.08458.pdf

Categories and Discriminators

Imperial College London

- **For the model independent resonance search:** Use both low and high mass analysis
- **Vector Leptoquark search:** Use only high mass analysis.
- **MSSM search:** Use high mass analysis combined with the CMS SM H→ττ analysis. http://cds.cern.ch/record/2725590?ln=en

High Mass Analysis ($m_\phi \geq 250$ GeV)



	No b-tag	b-tag
$\tau\tau \to e\mu$	Low-D_ζ Medium-D_ζ High-D_ζ	Low-D_ζ Medium-D_ζ High-D_ζ
$\tau\tau \to e\tau_h$	Loose-m_T Tight-m_T	Loose-m_T Tight-m_T
$\tau\tau \to \mu\tau_h$	Loose-m_T Tight-m_T	Loose-m_T Tight-m_T
$\tau\tau \to \tau_h\tau_h$	☐	☐
$t\bar{t}(e\mu)$		▨

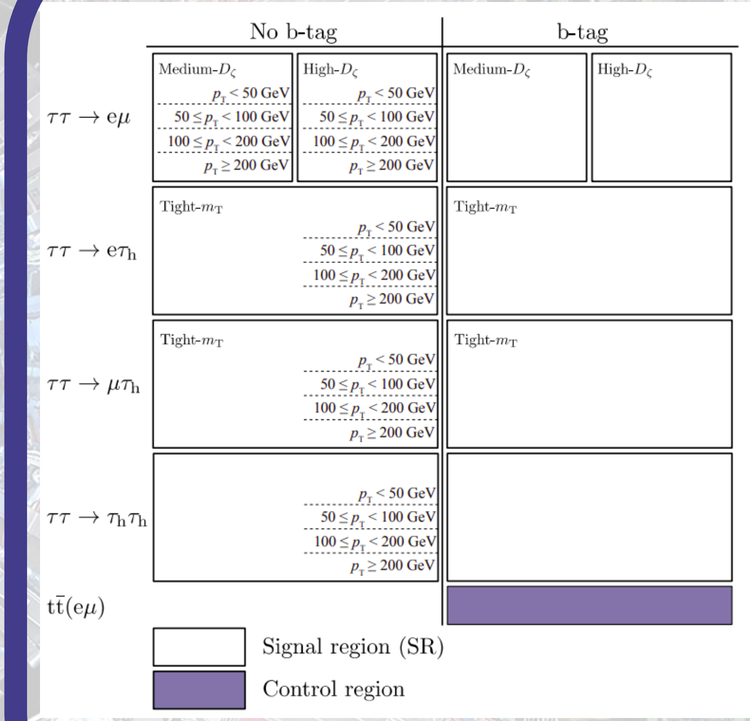
☐ Signal region (SR)

▨ Control region

Discriminator:

$$m_T^{tot} = [m_T(\tau_1,\tau_2)^2 + m_T(\tau_1,E_T^{miss})^2 + m_T(\tau_1,E_T^{miss})^2]^{1/2}$$

Low Mass Analysis ($m_\phi < 250$ GeV)



	No b-tag		b-tag	
$\tau\tau \to e\mu$	Medium-D_ζ $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV	High-D_ζ $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV	Medium-D_ζ	High-D_ζ
$\tau\tau \to e\tau_h$	Tight-m_T $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		Tight-m_T	
$\tau\tau \to \mu\tau_h$	Tight-m_T $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		Tight-m_T	
$\tau\tau \to \tau_h\tau_h$	$p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		☐	
$t\bar{t}(e\mu)$			▨	

☐ Signal region (SR)

▨ Control region

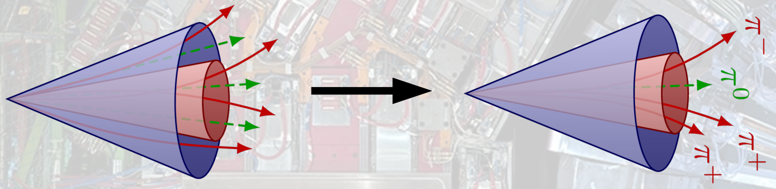
Discriminator:
$m_{\tau\tau}$ **(SVFit mass)**

https://iopscience.iop.org/article/10.1088/1742-6596/513/2/022035/pdf

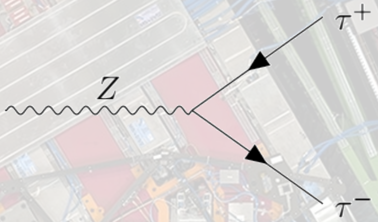
Background Modelling

- The most dominant backgrounds are from genuine di-τ pairs or jet→τ_h misidentification.

Jets Faking Hadronic Taus

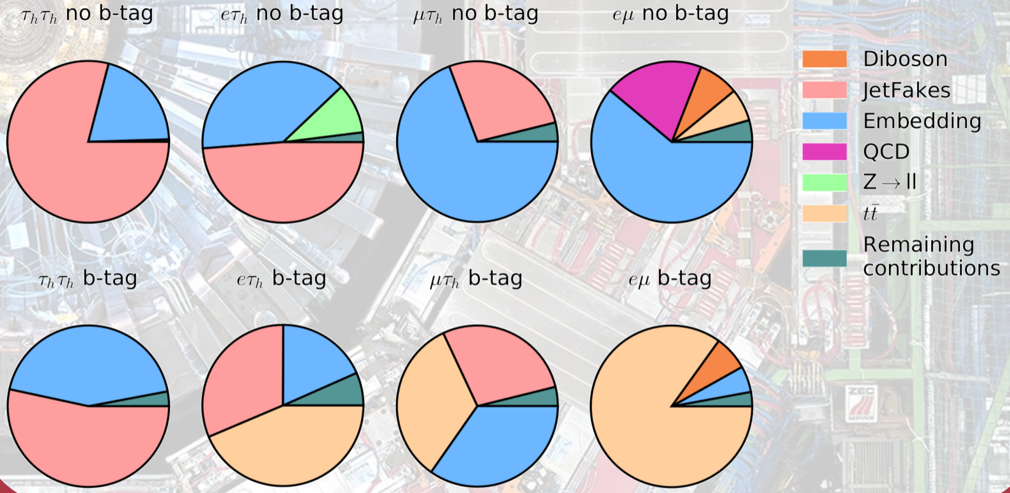


Genuine Di-tau Production



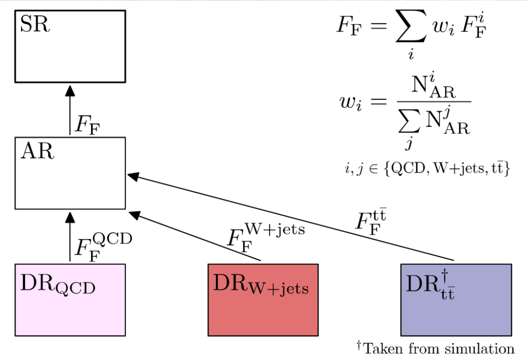
- Other backgrounds from processes with <2τ e.g. diboson, $t\bar{t}$, Z→ll.
- Di-τ (embedding method), jet→τ_h backgrounds (F_F method), and QCD (same sign τ pairs) are estimated from data-driven methods. ~90% of background modelling.
- Smaller remaining backgrounds estimated from MC.

Background Compositions



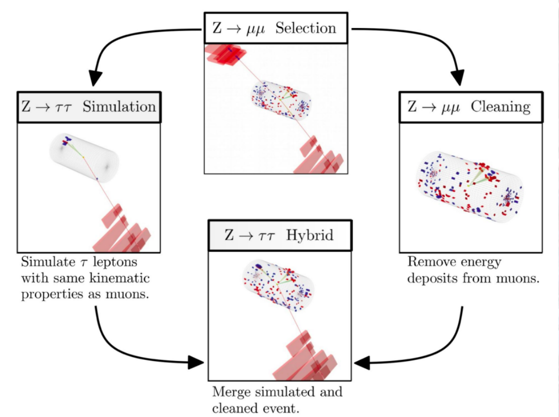
Data-Driven Methods

F_F Method



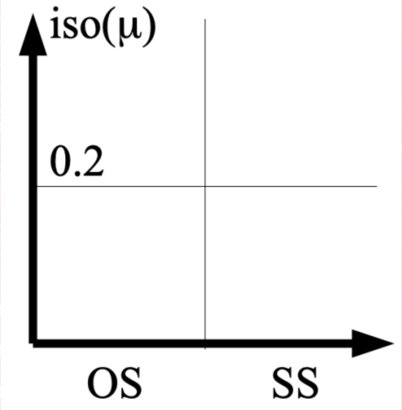
- Estimates all backgrounds with jets faking hadronic taus ($j \rightarrow \tau_h$).
- Calculate weight (F_F) as the ratio of two Tau ID working points from fake τ_h enriched sideband regions.
- Apply weights to data passing alternative working point.

Embedding Method



- Estimates backgrounds with real di-τ pairs, mainly $Z \rightarrow \tau\tau$, but a smaller contribution from $t\bar{t}$ and diboson.
- Replace muons selected in data with simulated τ lepton decays, utilising lepton flavour universality.

QCD Model

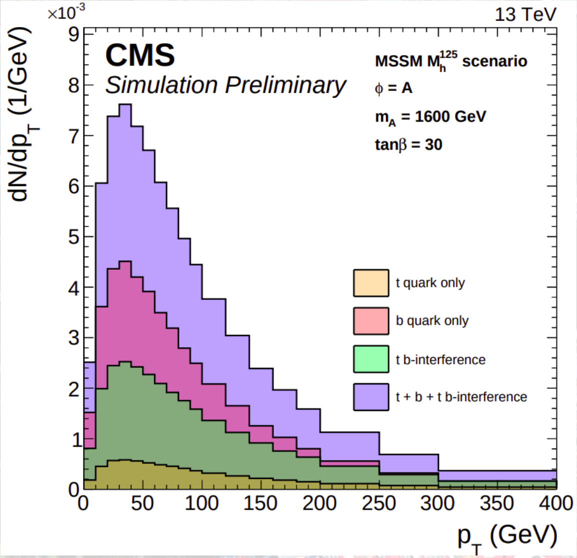


- Estimated (only in the eμ) in a sideband with same sign eμ pair.
- Use correction based off muon isolation.

Signal Modelling

Gluon fusion and b associated production signals

- Masses between 60 GeV and 3.5 TeV.
- p_T reweighting to separate gluon fusion loop.

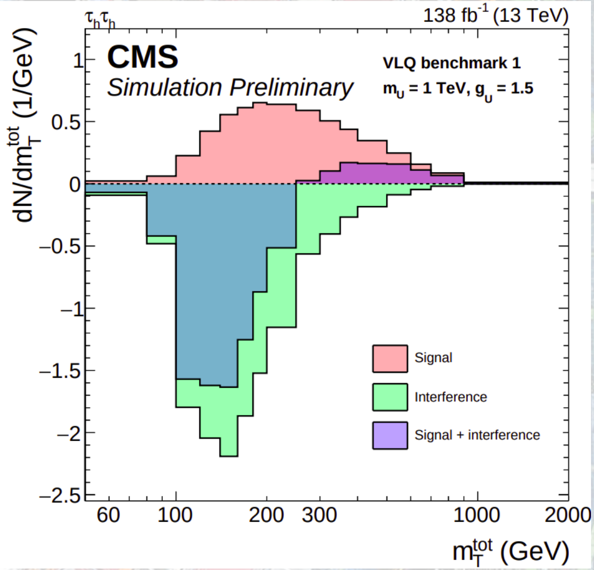


Vector leptoquark signal

- Look for U_1 vector leptoquark through t-channel process with mass between 1 and 5 TeV.

$$\mathcal{L}_U = \frac{g_U}{\sqrt{2}} U^\mu \left[\beta_L^{i\alpha}(\bar{q}_L^i \gamma_\mu l_L^\alpha) + \beta_R^{i\alpha}(\bar{d}_R^i \gamma_\mu e_R^\alpha)\right] + \text{h.c.} \quad \beta_L = \begin{pmatrix} 0 & 0 & \beta_L^{d\tau} \\ 0 & \beta_L^{s\mu} & \beta_L^{s\tau} \\ 0 & \beta_L^{b\mu} & \beta_L^{b\tau} \end{pmatrix}, \quad \beta_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_R^{b\tau} \end{pmatrix}$$

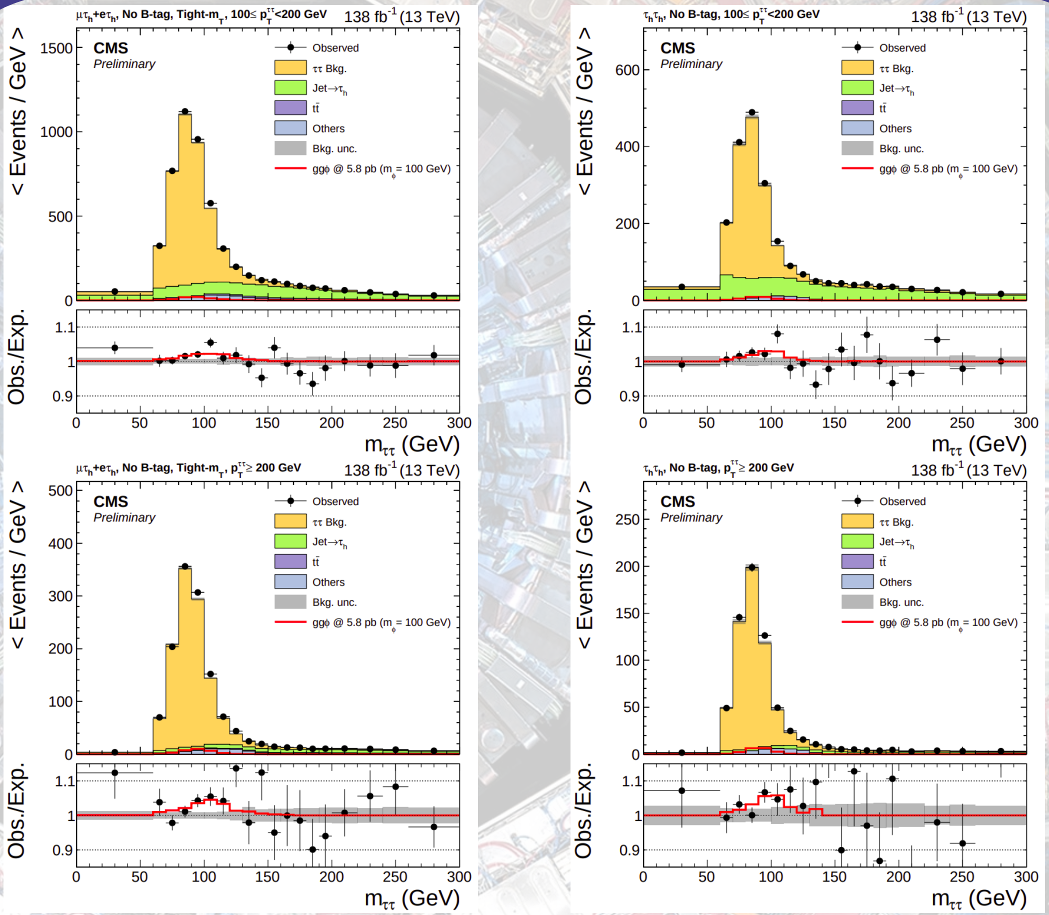
- Two benchmark (BM) scenarios considered:
  - VLQ BM 1: $\beta_R^{b\tau}=0$ - no RH couplings
  - VLQ BM 2: $\beta_R^{b\tau}=-1$ - Pati-Salam-like leptoquark
- Other matrix parameters set to best fit of b anomalies.
- Interference with background also included.



Phenomenology follows https://arxiv.org/pdf/2103.16558.pdf

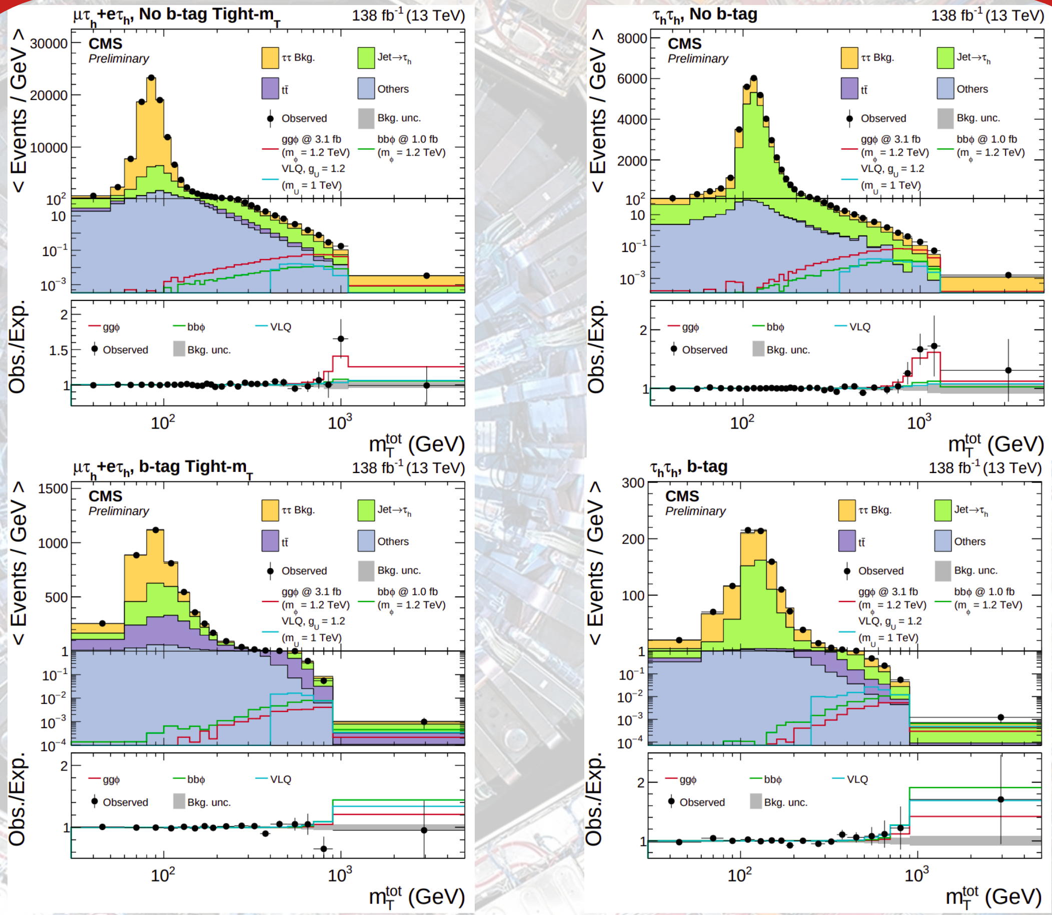
Results – Postfit Low Mass

- Most sensitive categories shown, with all years and semi-leptonic tau decay channels combined.
- Small excess observed in $m_{\tau\tau}$ bins around 100 GeV.
- The best fit signal strengths for ggϕ is shown.
- The excess is fit well by a **ggϕ resonance of mass 100 GeV**.



Results – Postfit High Mass

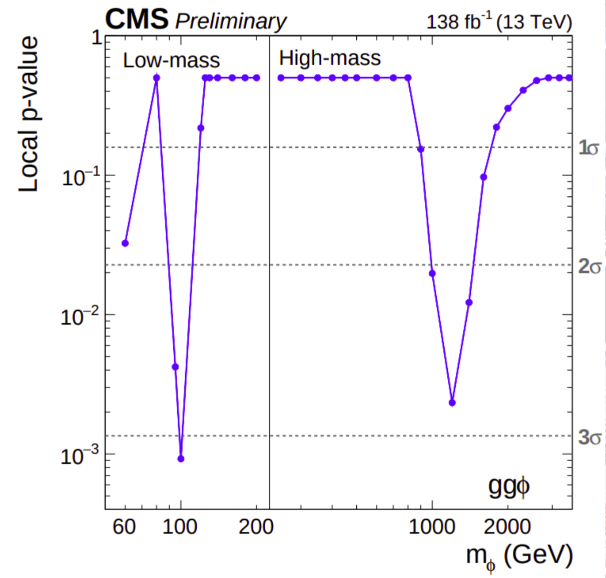
- Most sensitive categories shown, with all years and semi-leptonic tau decay channels combined.
- Small excess observed in high m_T^{tot} bins.
- The best fit signal strengths for ggϕ, bbϕ and VLQ BM 1 are drawn.
- The excess is fit well by a **ggϕ resonance of mass 1.2 TeV**.
- The excess is **not** explained by the VLQ model, as b-tag categories are more sensitive.



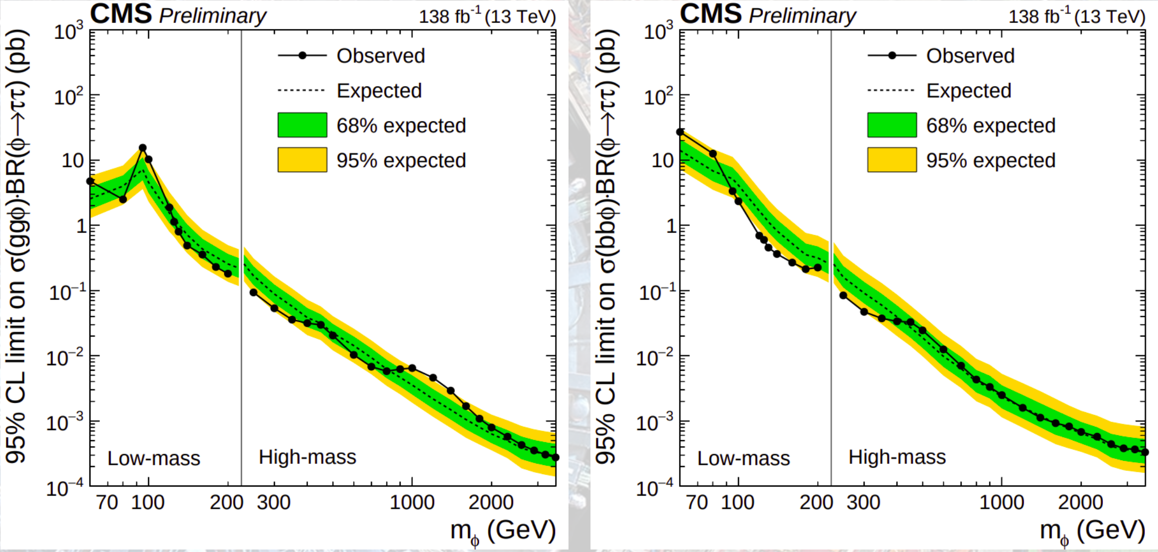
Results – Model Independent Limits and p-value Scan

- 95% CL limits are placed on gluon fusion production and production in associated with b-quarks.

p-value Scan for Gluon Fusion



Model Independent Limits

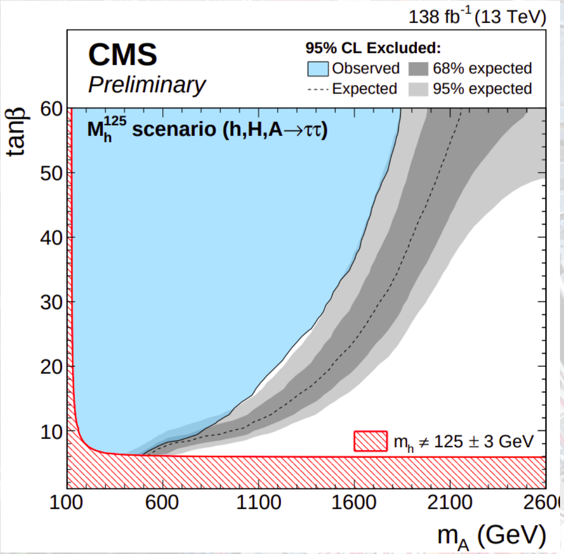


- The excesses observed in gluon fusion production have the below local (global) significances:
  - **At 100 GeV: 3.1σ (2.7σ)**
  - **At 1.2 TeV: 2.8σ (2.4σ)**

Results – Model dependent Limits

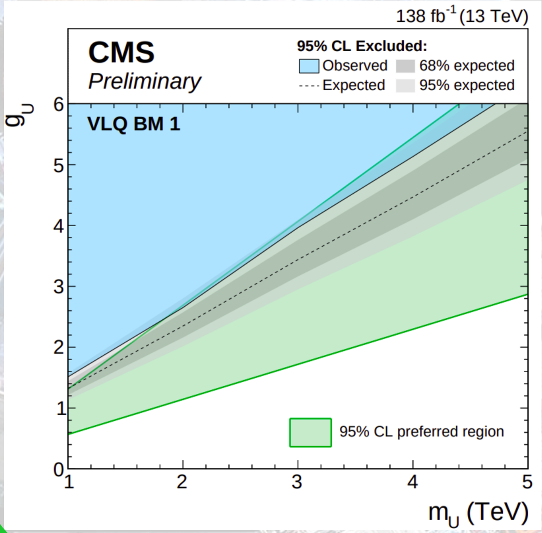
MSSM

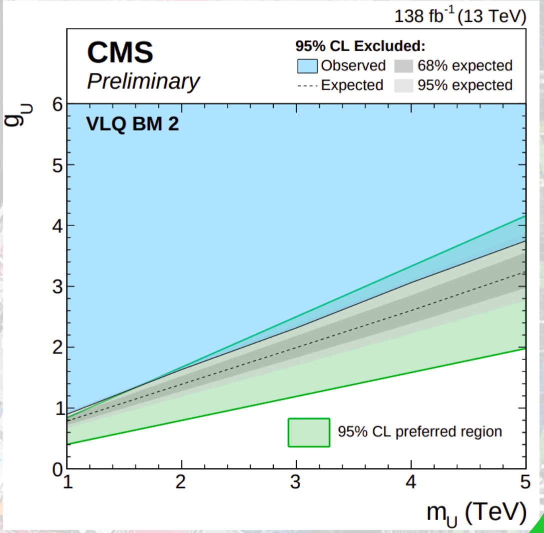
- Model dependent 95% exclusion contours shown for MSSM Scenario M_h^{125}.
- Red hatched band – theory uncertainty.



Vector Leptoquark

- The green shaded band represents the 95% CL preferred region for a vector leptoquark to explain the B-anomalies.
- Analysis sensitive to vector leptoquark that could explain the B-anomalies.

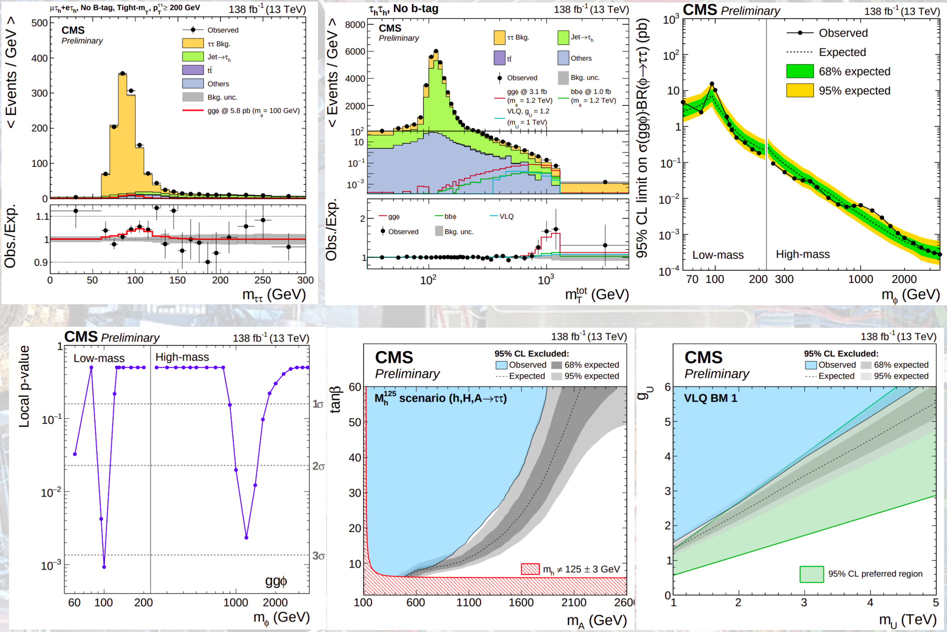




Results – Conclusions

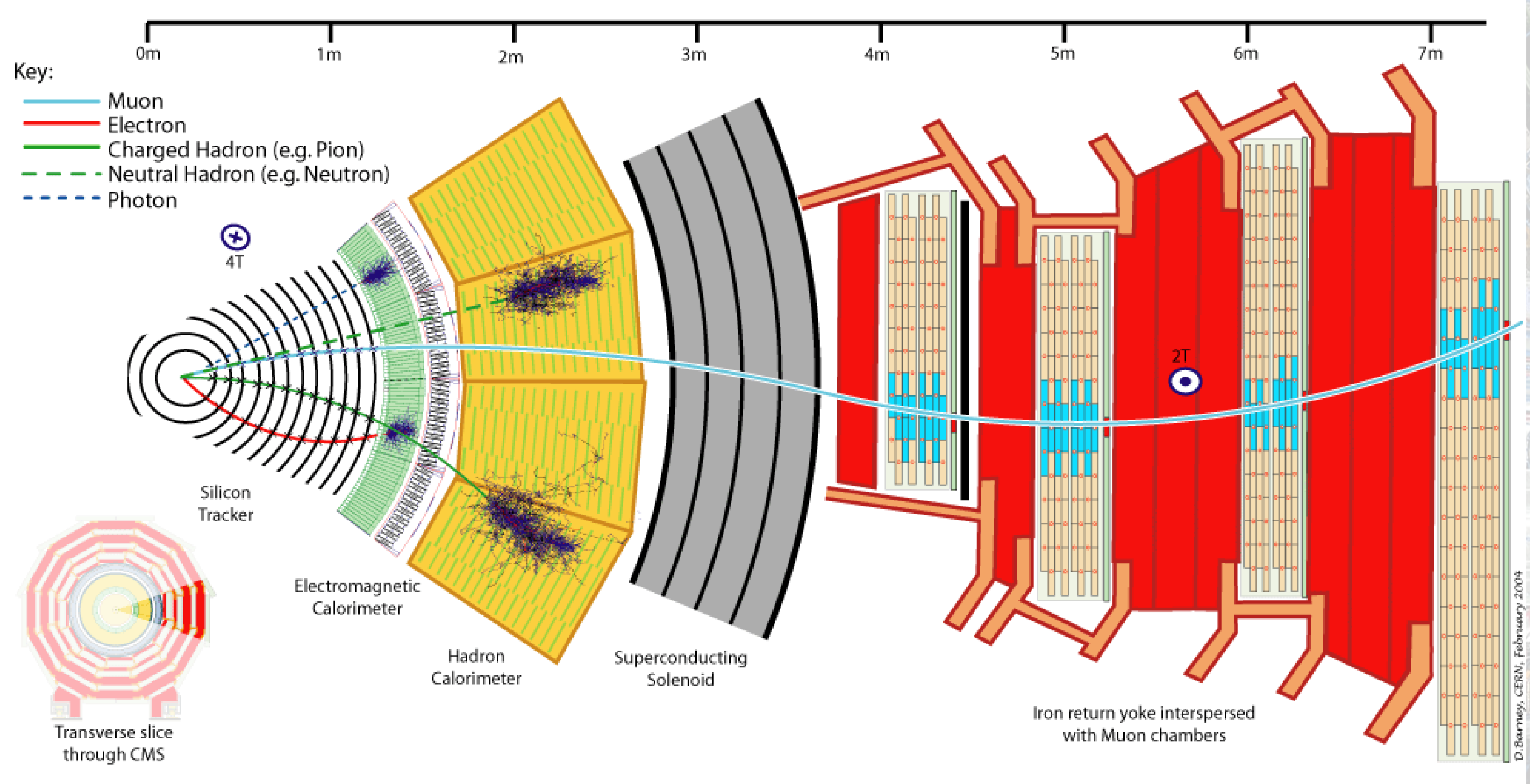
- Two local excess observed for gluon fusion production but no signals with global significances larger than 3σ.
  - m_ϕ = 100 GeV: Local (global) significance 3.1σ (2.7σ).
  - m_ϕ = 1.2 TeV: Local (global) significance 2.8σ (2.4σ).
- The observed b associated production lies for all mass points with 2σ of the expectation.
- Strongest constraints placed on the MSSM Higgs sector.
- Excess not compatible with either leptoquark model considered as you would expect more events in b-tag categories.
- Analysis sensitive to leptoquarks that could explain the B-anomalies.
- Presentated at Moriond 2022, link to conference note **here**.

Summary of Results



Backup

CMS Experiment



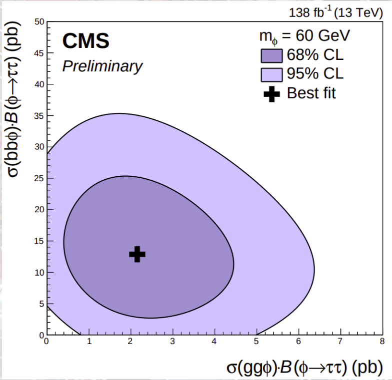
Systematics

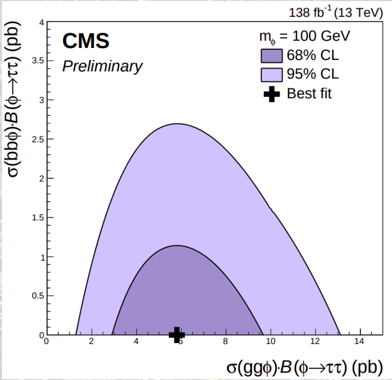
Imperial College London

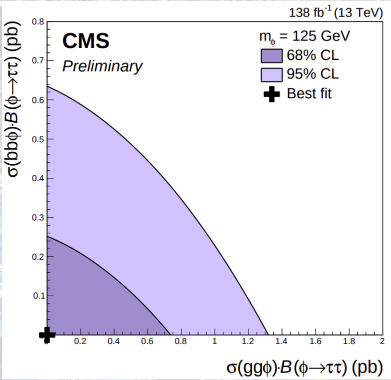
- Luminosity
- e,μ id, iso and trigger efficiency and energy scales (for e only)
- Hadronic tau ID and trigger efficiency and energy scales
- b-tagging efficiency and mis-tag
- JES / JER
- Background cross sections
- l → τ_h fake rates and energy scales
- Embedding normalization due to selection efficiency of double muon trigger
- Embedding p_T/mass shape from Z→μμ closures
- Fake-factor uncertainties (statistical and systematic)
- QCD (eμ) OS/SS statistical and systematic uncertainties
- Prefiring
- Embedded MET energy scale and resolution
- MET Recoil corrections
- DY mass and pT reweighting contamination in embedded samples
- Top quark pT reweighting
- Bin-by-bin uncertainties
- Signal theory uncertainties

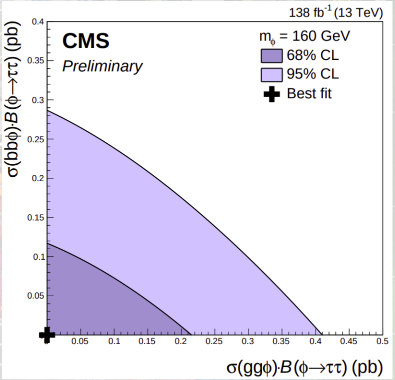
2D Likelihood Scans

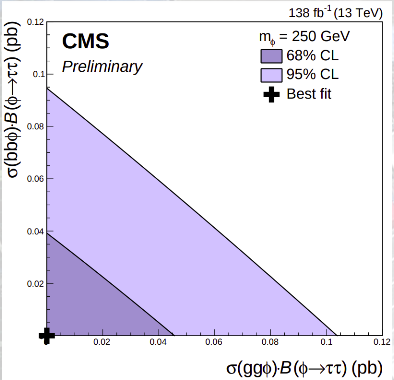
- 2D likelihood scans of gluon fusion and b associated production cross sections times di-tau branching fractions.

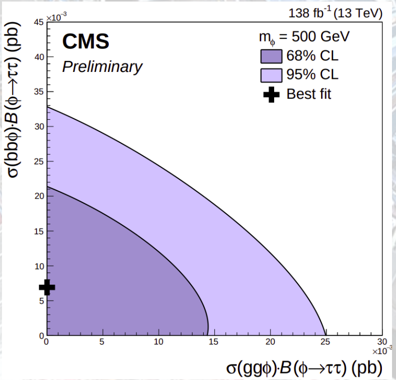


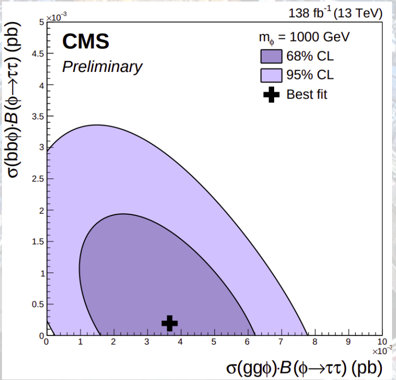


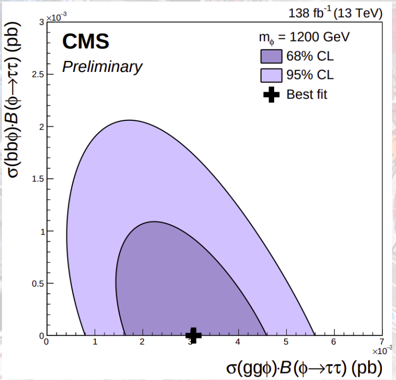


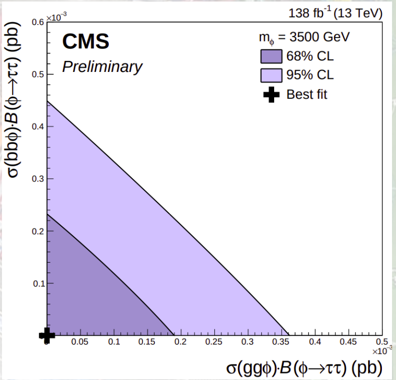






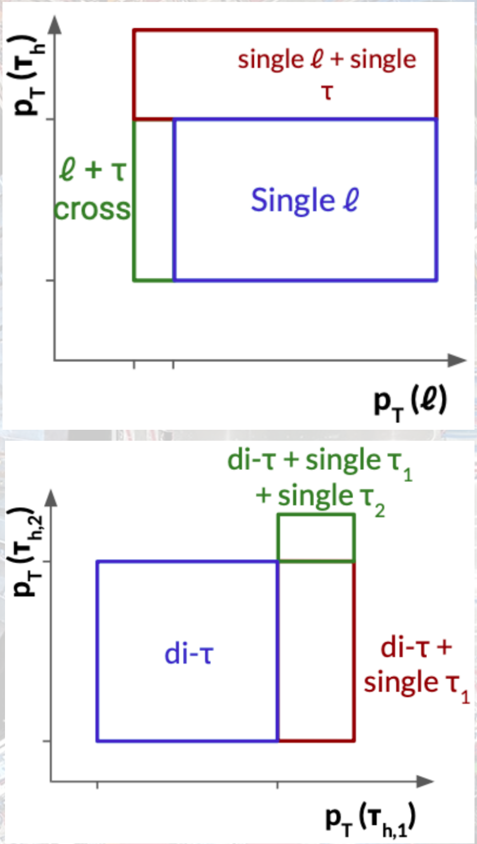






Triggers

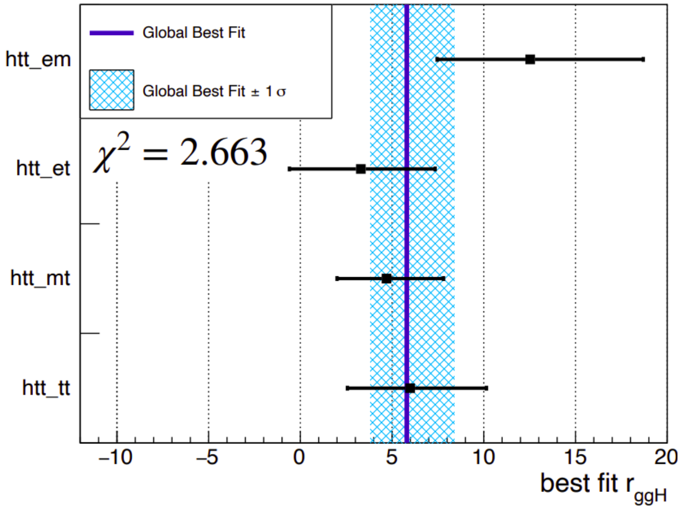
- Trigger Selections:
  - $\tau_h\tau_h$: single-tau OR double-tau
  - $e\tau_h$: single-electron OR electron-τ_h cross OR single-tau
  - $\mu\tau_h$: single-muon OR muon-τ_h cross OR single-tau
  - eμ: Muon-electron cross
- Thresholds:
  - eμ: 2016/2017/2018: e(12) & μ(23), e(23) & μ(8)
  - $e\tau_h$: 2016: e(25), e(24) & τ_h(20/30), τ_h(120)
    2017: e(27), e(32), e(35), e(24) & τ_h(30), τ_h(180)
    2018: e(32), e(35), e(24) & τ_h(30), τ_h(180)
  - $\mu\tau_h$: 2016: μ(22), μ(19) & τ_h(20), τ_h(120)
    2017: μ(24), μ(27), μ(20) & τ_h(27), τ_h(180)
    2018: μ(24), μ(27), μ(20) & τ_h(27), τ_h(180)
  - $\tau_h\tau_h$: 2016: τ_h(35) & τ_h(35), τ_h(120)
    2017: τ_h(35) & τ_h(35), τ_h(40) & τ_h(40), τ_h(120)
    2018: τ_h(35) & τ_h(35), τ_h(40) & τ_h(40), τ_h(120)



Low Mass Compatibility Plots

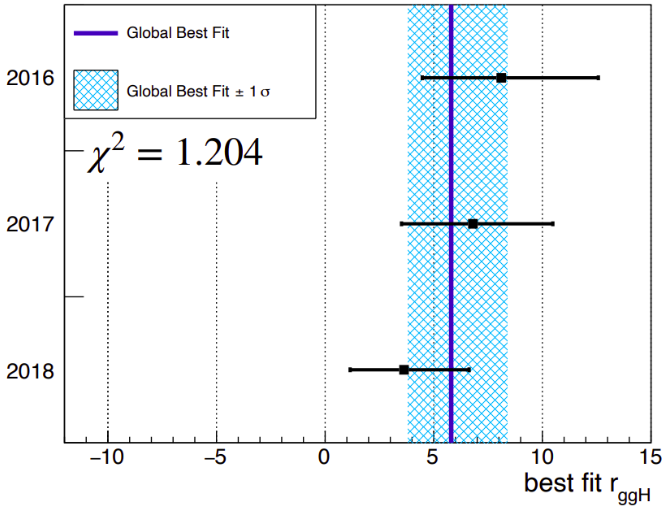
- r_{ggH} = σ(ggϕ) X BR(ϕ→ττ) (pb)
- 100 GeV gluon fusion resonance shown.

m_{ϕ}=100 GeV



p-value=0.50

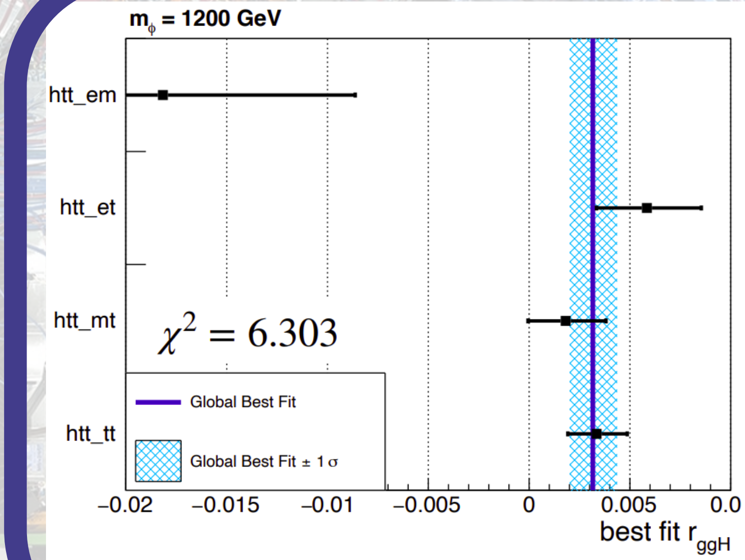
m_{ϕ}=100 GeV



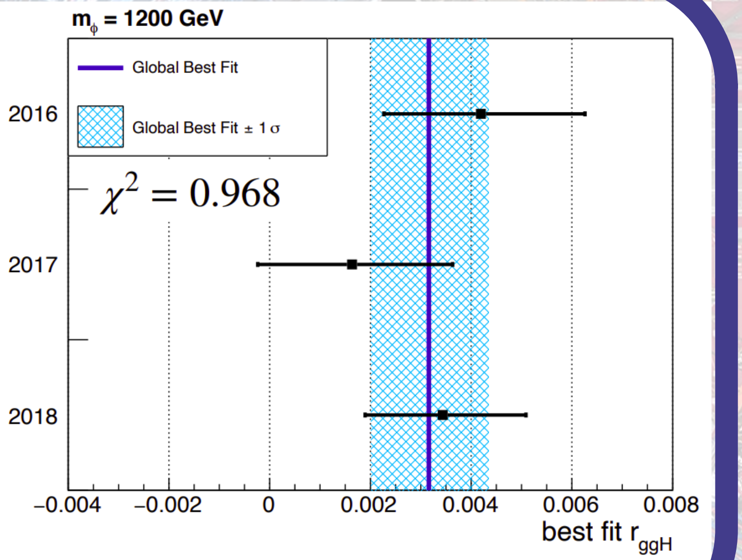
p-value=0.58

High Mass Compatibility Plots

- $r_{ggH} = \sigma(gg\phi) \times BR(\phi \rightarrow \tau\tau)$ (pb)
- 1.2 TeV gluon fusion resonance shown.



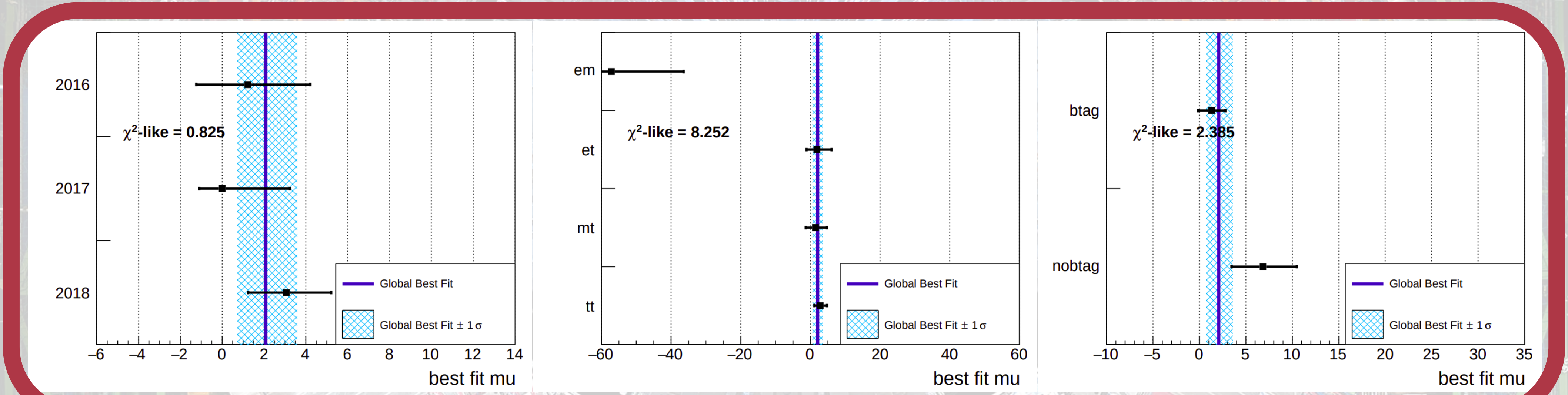
p-value=0.11



p-value=0.64

VLQ Compatibility Plots

- μ(mu) = g_U^4. This is defined so to allow the lower uncertainty fit to go negative in case of no crossing.
- VLQ BM 1 at 1 TeV shown.



MSSM Limit Setting

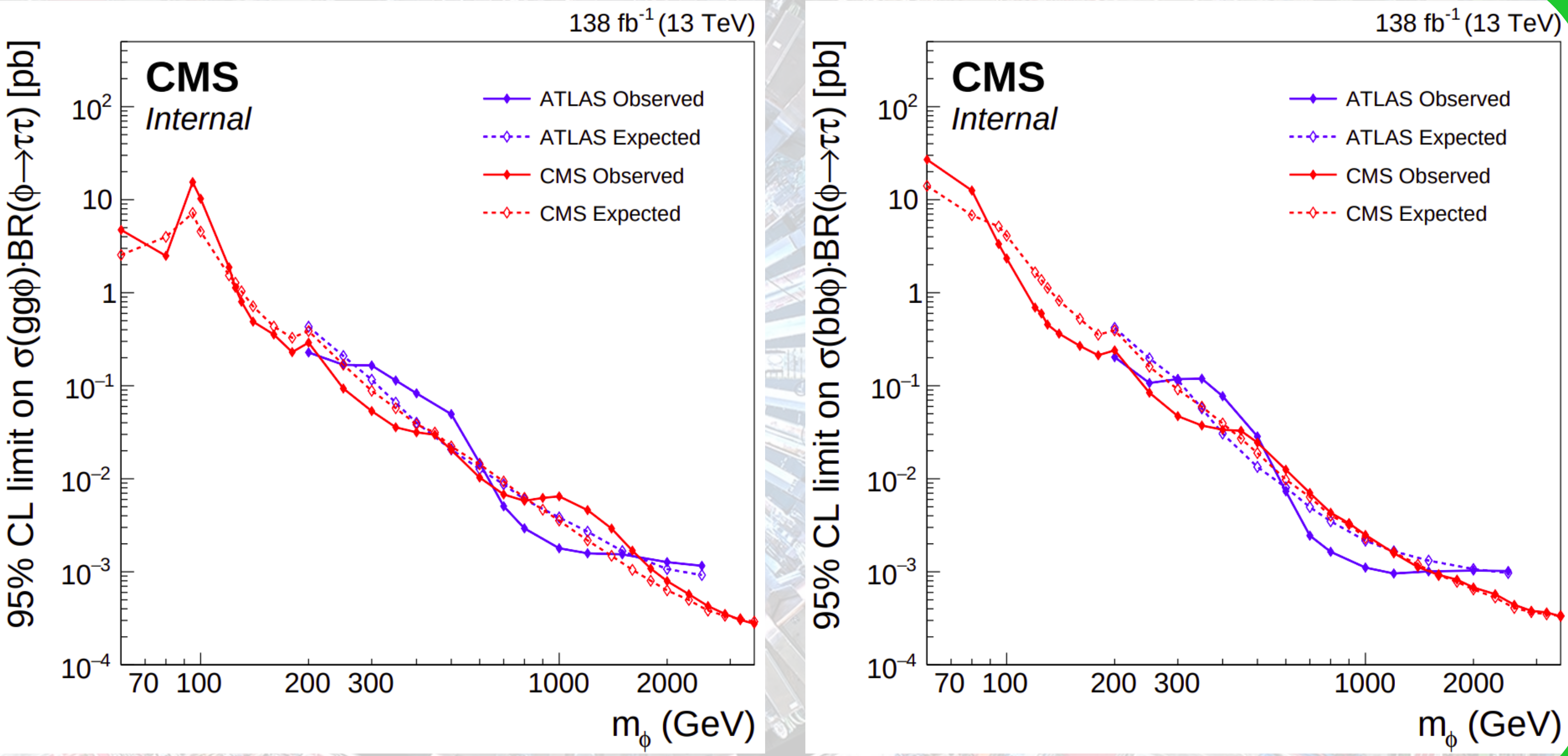
- MSSM signal model composed as linear combination of templates for h, H, and A bosons.
- SM and BSM contribution BSM templates scaled to XS and BR prediction from benchmark scenario.
- Hypothesis Test: For each considered parameter point in the m_A, tanβ plane of the considered benchmark scenario perform hypothesis test for BSM vs. SM model prediction with the LHC test statistics.

$$q_\mu = -\,2\frac{\mathscr{L}(\mathrm{data}\,|\,\mu, \hat{\theta}_\mu)}{\mathscr{L}(\mathrm{data}\,|\,\hat{\mu}, \hat{\theta}_{\hat{\mu}})}$$

$$\mathscr{L}(\mathrm{data}\,|\,\mu) = \mathscr{L}\left(\mathrm{data}\,|\,\mu\left((s_{\mathrm{h}} - S_{\mathrm{SM}}) + S_{\mathrm{H}} + S_{\mathrm{A}}\right) + S_{\mathrm{SM}} + b\right)$$

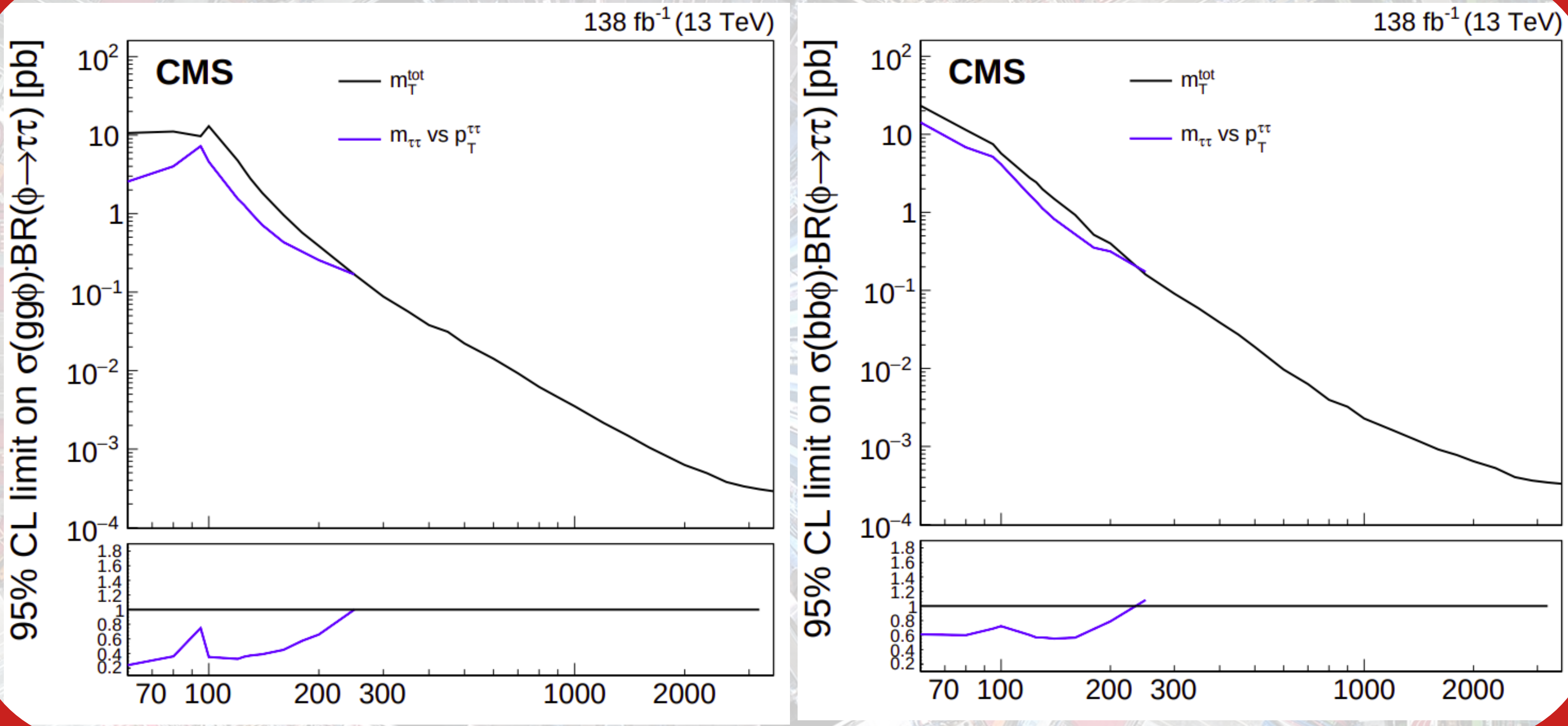
ATLAS Comparison

- Comparison of ATLAS and CMS model independent limits.
- ATLAS analysis searches for a mass range of 200 GeV to 2.2 TeV.



Expected Limit Sensitivity Comparison

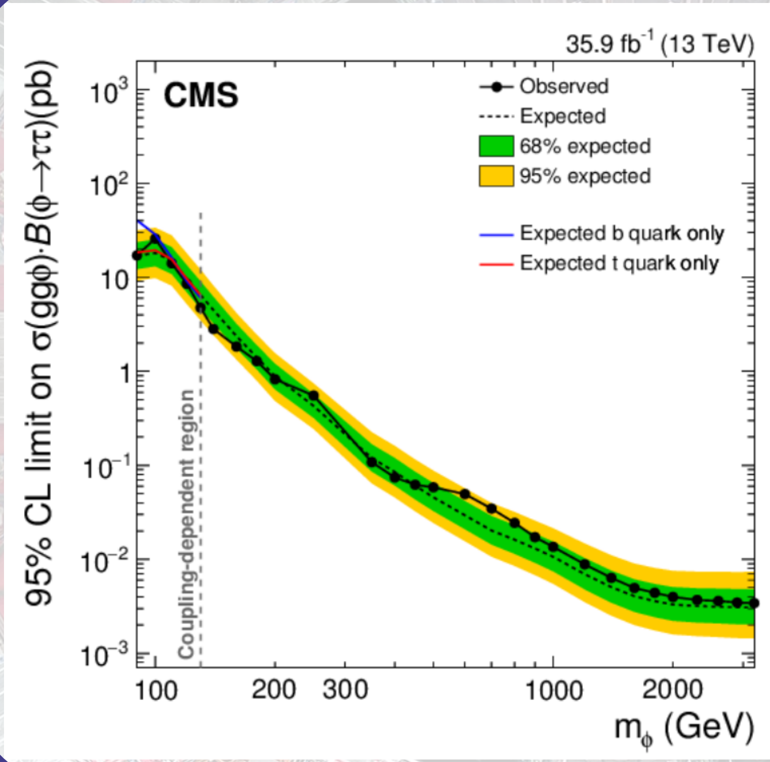
- Comparison of using m_T^{tot} as discriminator instead of the $m_{\tau\tau}$ vs $p_T^{\tau\tau}$.

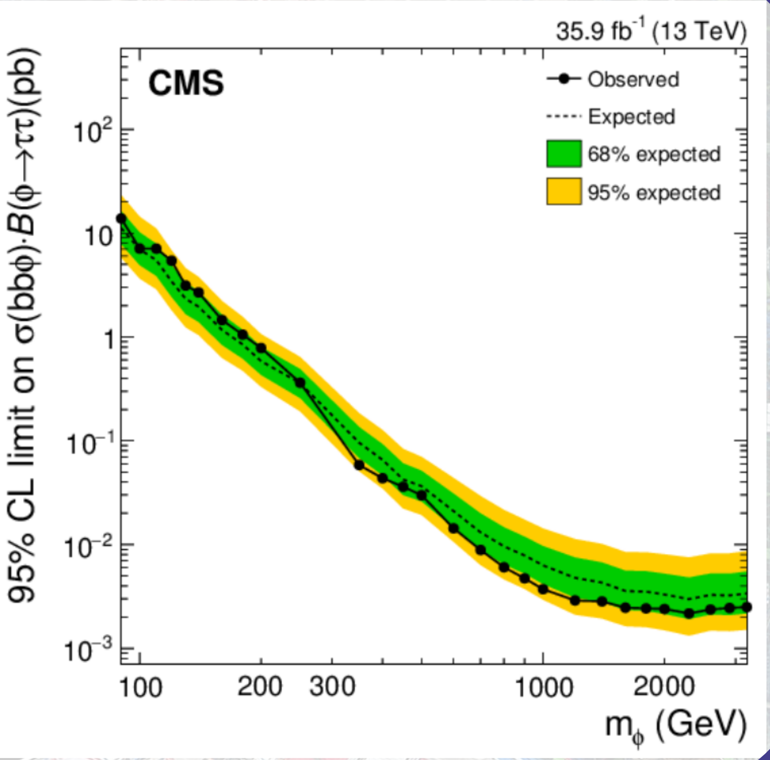


Results for Partial Run-2 Analysis

- Model independent limits from the CMS analysis performed on 2016 data.

https://arxiv.org/pdf/1803.06553.pdf

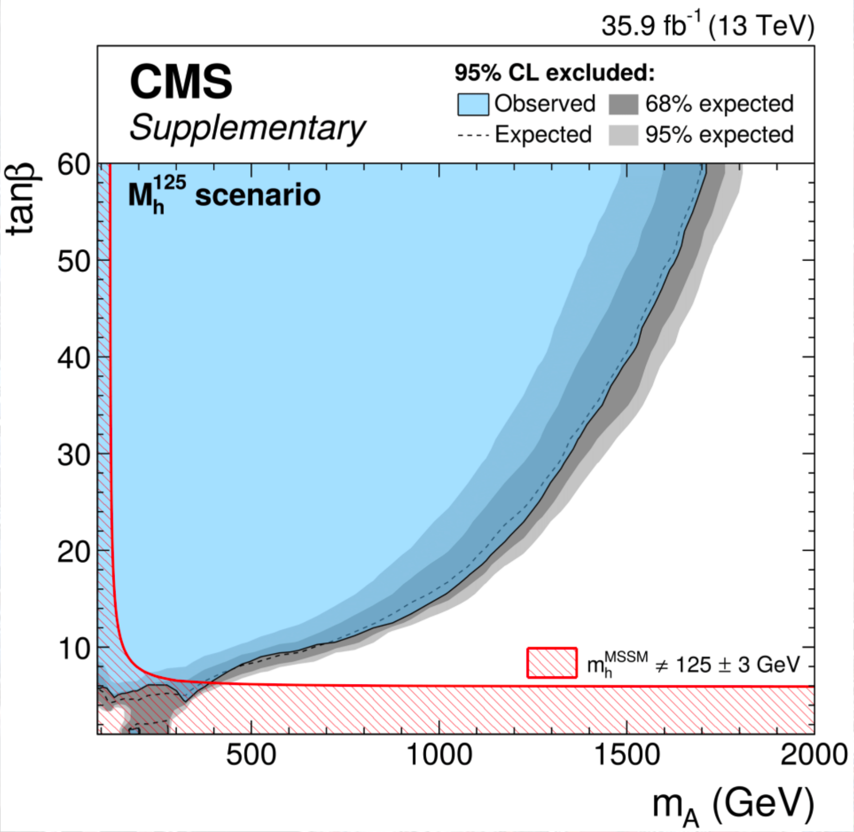




Results for Partial Run-2 Analysis

- Model dependent limits on the MSSM M_h^{125} scenario from the CMS analysis performed on 2016 data.

https://arxiv.org/pdf/1803.06553.pdf

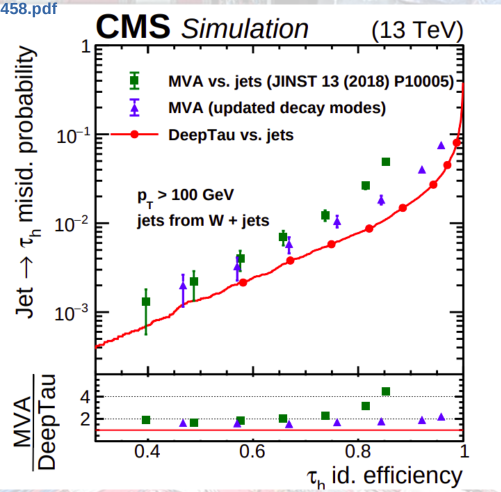


DeepTau Algorithm

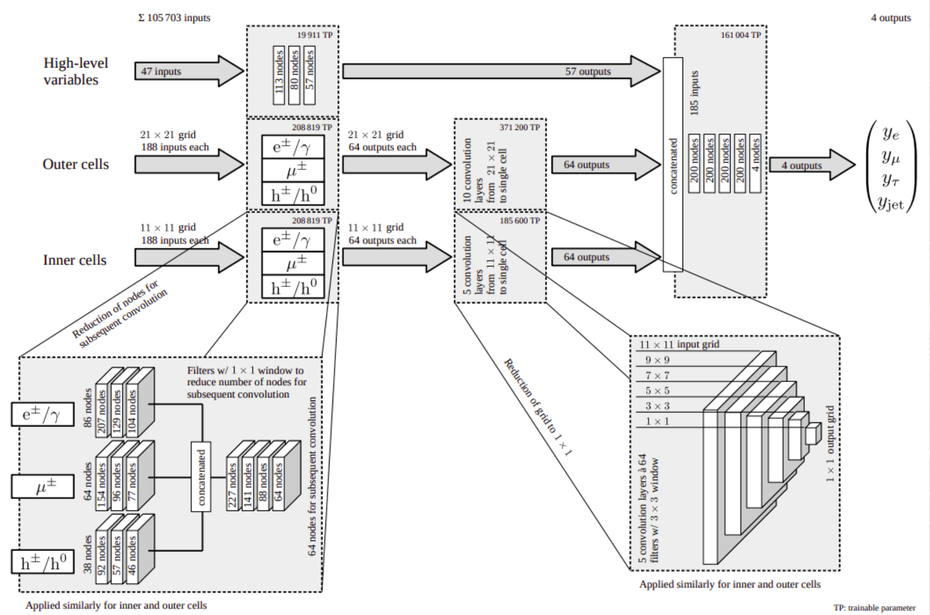
- DeepTau is a multiclass tau identification algorithm based on a convolutional deep neural network (DNN).
- DeepTau combines information from the high-level reconstructed tau features together with the low level information from the inner tracker, calorimeters and muon sub-detectors within the tau isolation cone.

https://arxiv.org/pdf/2201.08458.pdf

ROC Curve



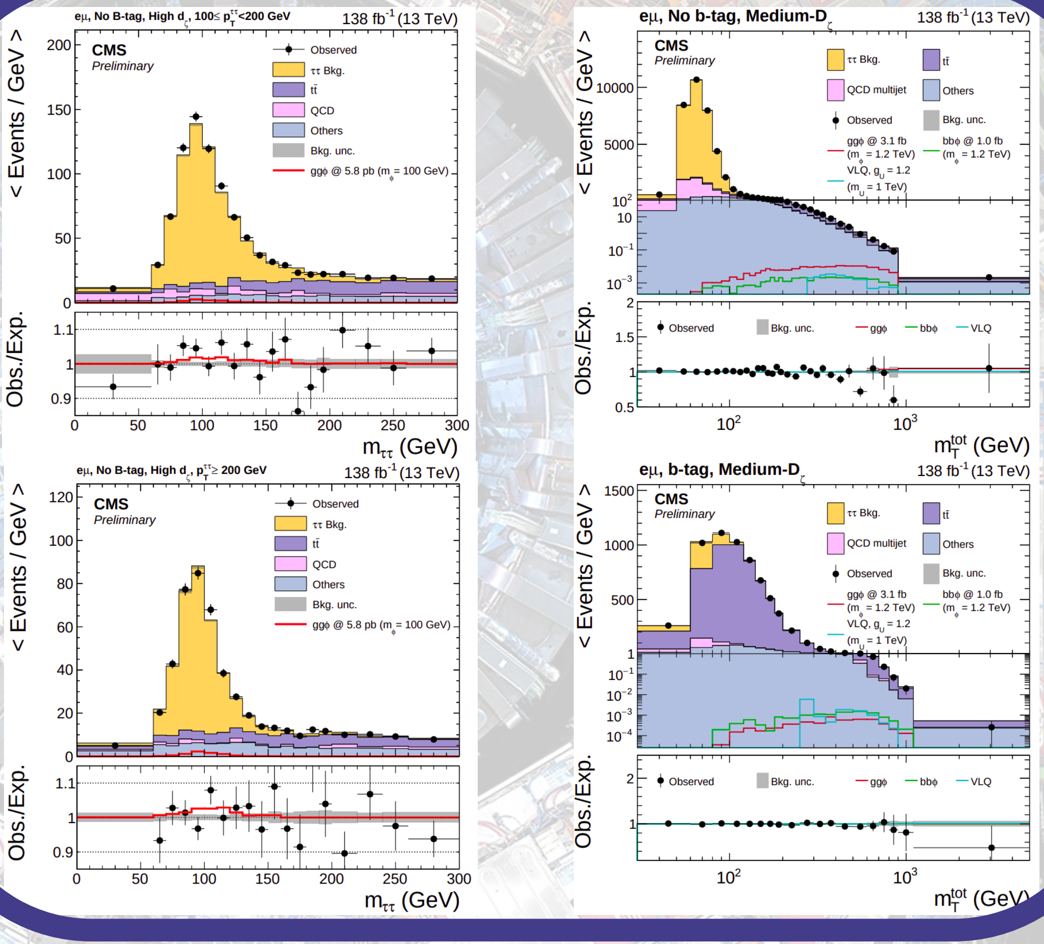
Architecture



https://arxiv.org/pdf/2201.08458.pdf

eμ Postfit Plots

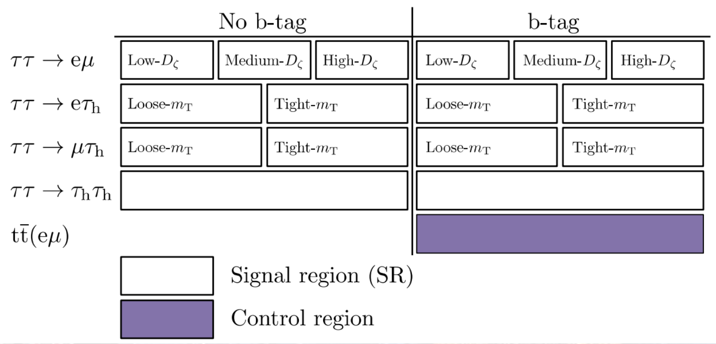
- Postfit plots for the most sensitive category in the eμ decay channel.



Variable Definitions

- Definitions of variables used in categorisation and discriminators.

High Mass Analysis ($m_\phi \geq 250$ GeV)



	No b-tag			b-tag		
$\tau\tau \rightarrow e\mu$	Low-D_ζ	Medium-D_ζ	High-D_ζ	Low-D_ζ	Medium-D_ζ	High-D_ζ
$\tau\tau \rightarrow e\tau_h$	Loose-m_T	Tight-m_T		Loose-m_T	Tight-m_T	
$\tau\tau \rightarrow \mu\tau_h$	Loose-m_T	Tight-m_T		Loose-m_T	Tight-m_T	
$\tau\tau \rightarrow \tau_h\tau_h$	☐			☐		
$t\bar{t}(e\mu)$				▨		

☐ Signal region (SR)

▨ Control region

Discriminator:

$m_T^{tot} = [m_T(\tau_1,\tau_2)^2 + m_T(\tau_1,E_T^{miss})^2 + m_T(\tau_1,E_T^{miss})^2]^{1/2}$

Variable definitions:

$D_\zeta = p_\zeta^{miss} - 0.85p_\zeta^{obs}$

$p_\zeta^{miss} = \mathbf{p}_\zeta^{\mathbf{miss}} \, . \, \boldsymbol{\zeta}$

$p_\zeta^{vis} = (\mathbf{p}_\zeta^{\mathbf{e}}+\mathbf{p}_\zeta^{\boldsymbol{\mu}}) \, . \, \boldsymbol{\zeta}$

$\boldsymbol{\zeta}$ is the vector that bisects $\mathbf{p}_\zeta^{\mathbf{e}}$ and $\mathbf{p}_\zeta^{\boldsymbol{\mu}}$

$m_T(A,B) = [2p_T^A p_T^B(1-\cos\Delta\phi^{AB})]^{1/2}$

Category definitions:

$t\bar{t}(e\mu)$: $D_\zeta < -35$ GeV

Low-D_ζ: $-35 < D_\zeta < 10$ GeV

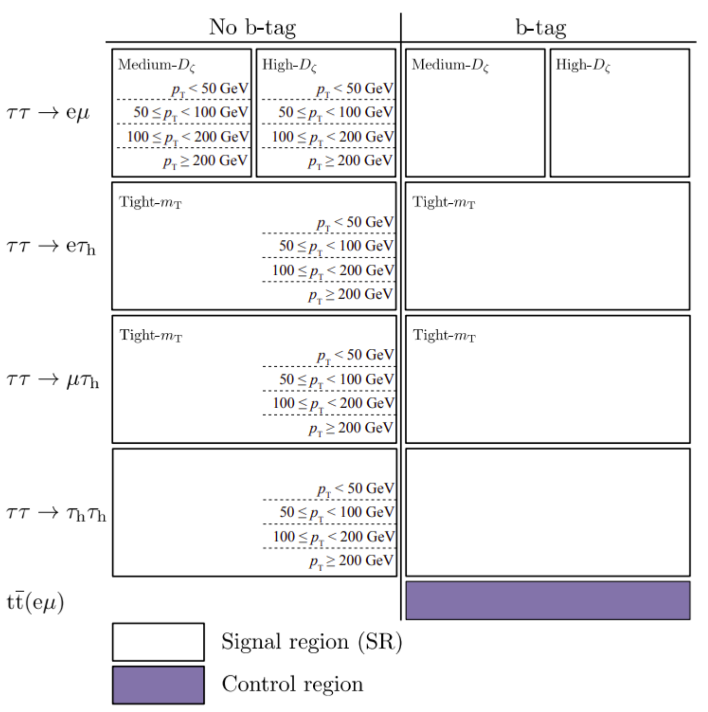
Medium-D_ζ: $10 < D_\zeta < 30$ GeV

High-D_ζ: $D_\zeta > 30$ GeV

Tight-m_T: $m_T(e/\mu,E_T^{miss}) < 40$ GeV

Loose-m_T: $40 < m_T(e/\mu,E_T^{miss}) < 70$ GeV

Low Mass Analysis ($m_\phi < 250$ GeV)



	No b-tag		b-tag	
$\tau\tau \rightarrow e\mu$	Medium-D_ζ $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV	High-D_ζ $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV	Medium-D_ζ	High-D_ζ
$\tau\tau \rightarrow e\tau_h$	Tight-m_T $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		Tight-m_T	
$\tau\tau \rightarrow \mu\tau_h$	Tight-m_T $p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		Tight-m_T	
$\tau\tau \rightarrow \tau_h\tau_h$	$p_T < 50$ GeV $50 \leq p_T < 100$ GeV $100 \leq p_T < 200$ GeV $p_T \geq 200$ GeV		☐	
$t\bar{t}(e\mu)$			▨	

☐ Signal region (SR)

▨ Control region

Discriminator:

$m_{\tau\tau}$ (SVFit mass)

https://iopscience.iop.org/article/10.1088/1742-6596/513/2/022035/pdf