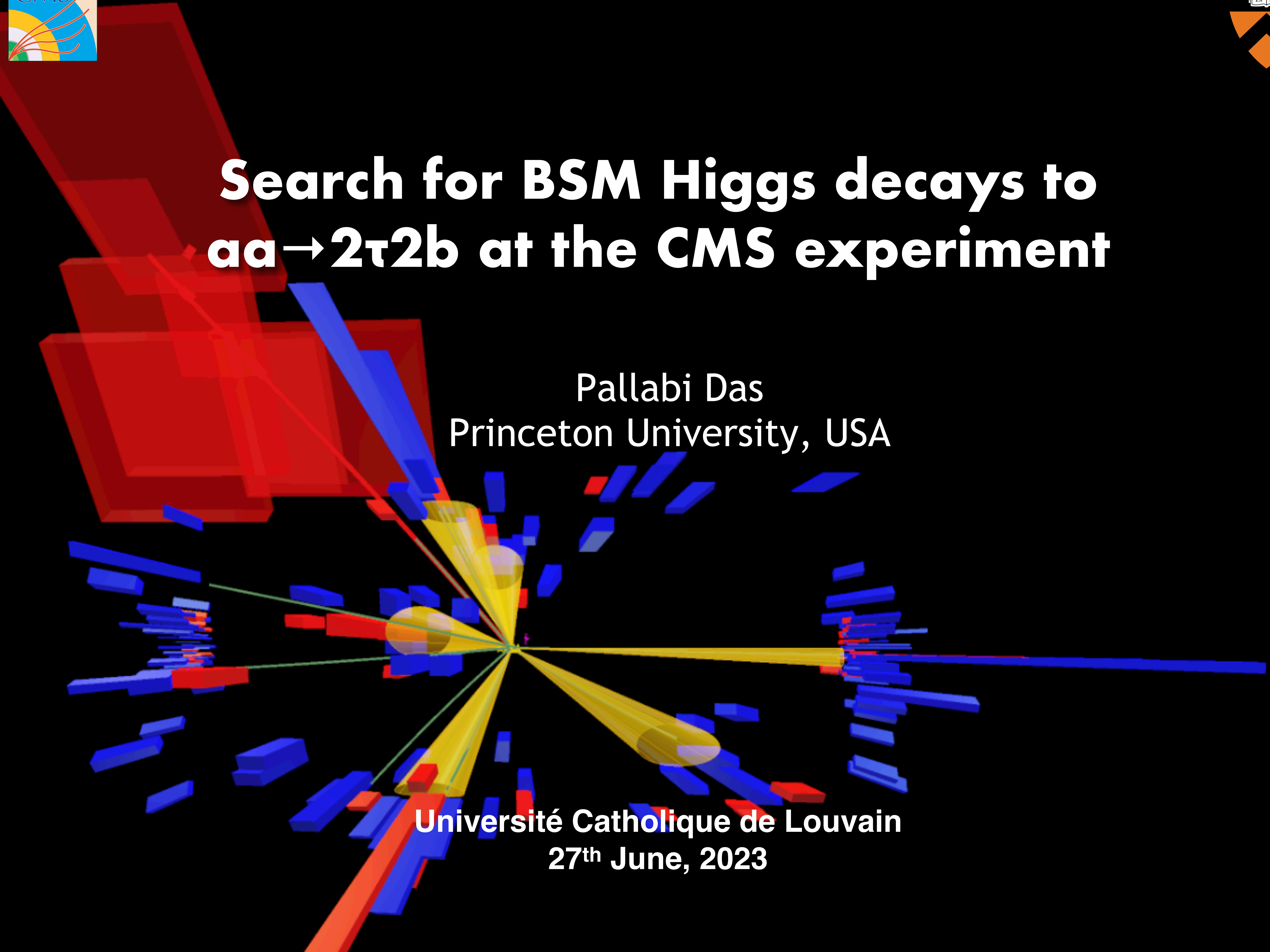


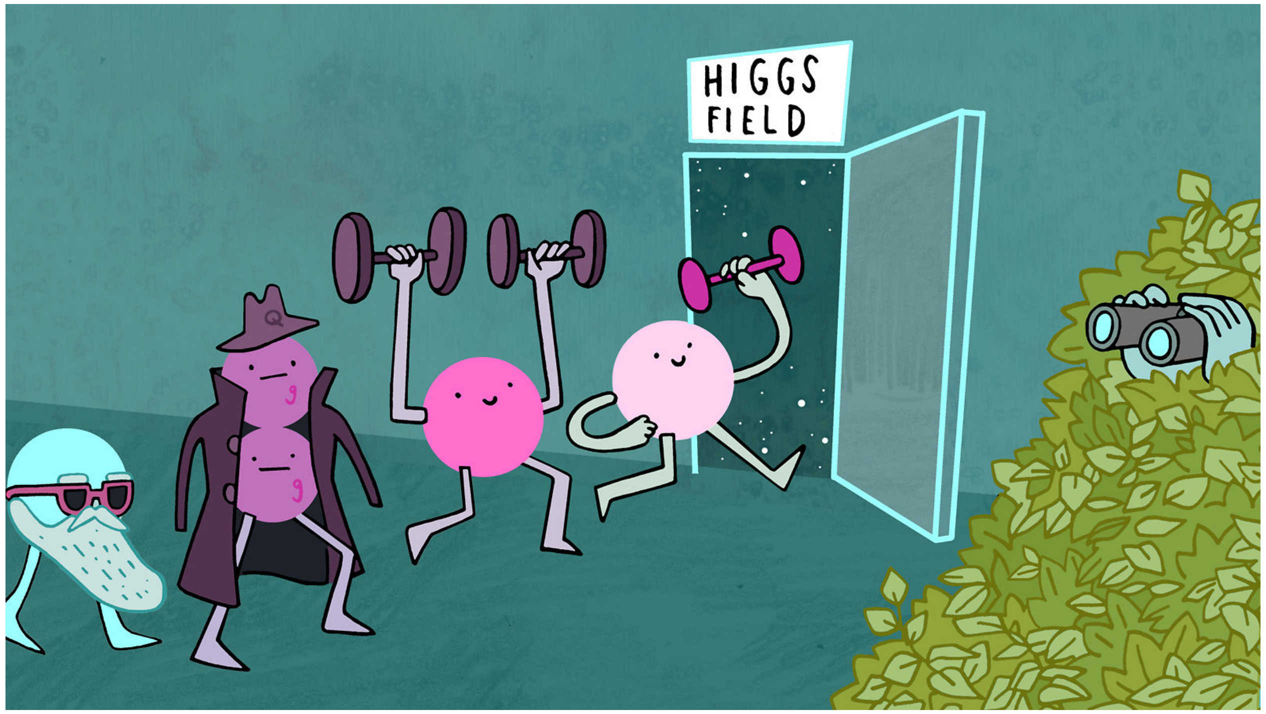
Search for BSM Higgs decays to $aa \to 2\tau 2b$ at the CMS experiment

Pallabi Das
Princeton University, USA

Université Catholique de Louvain
27th June, 2023

Contents

- Introduction
  - Motivation
  - LHC and CMS experiment
  - Theoretical models
- Analysis
  - Methodology
  - Object and event selection
  - DNN categorisation
  - Backgrounds
  - Uncertainties
- Results
  - Signal region distributions
  - Limits on exotic Higgs BR
- Summary



INTRODUCTION

Motivation

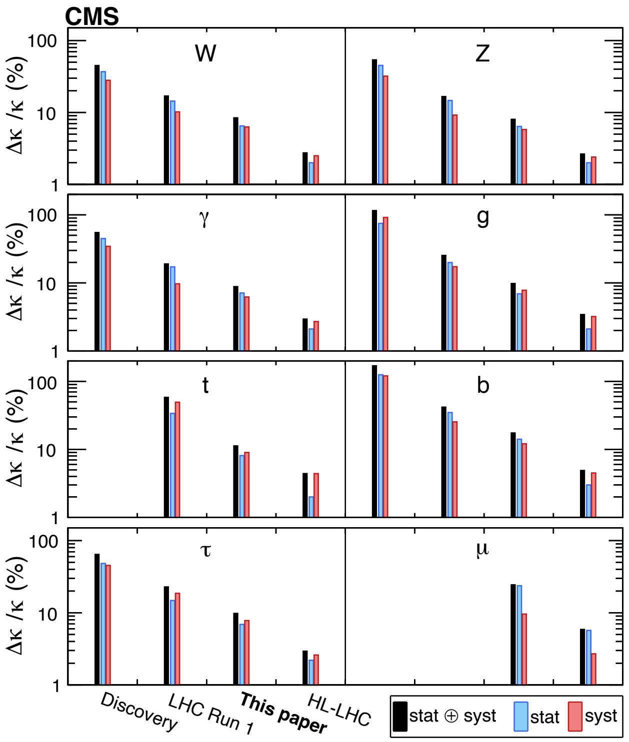
The discovery of the Higgs boson 10 years ago [1, 2] established the theory of the SM

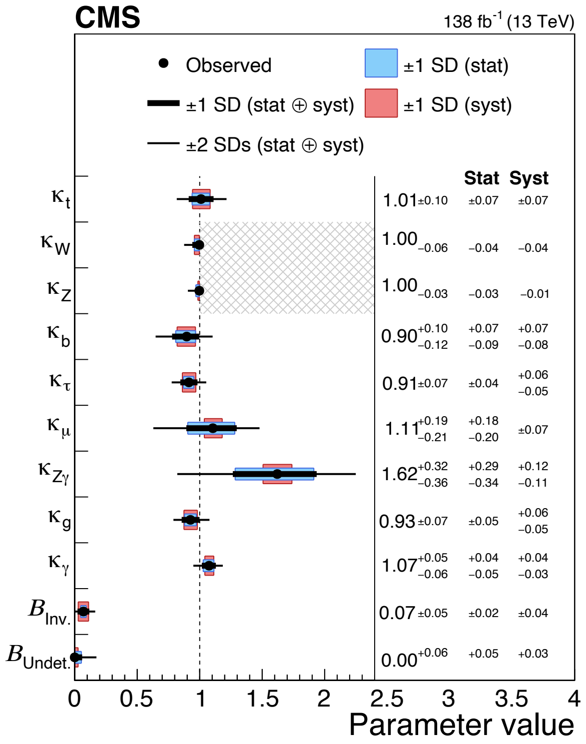
→ ***But many questions remain!***

- Several BSM theories which can explain **Dark Matter** origin, **Hierarchy Problem**, etc. and also predict a **Higgs Resonance**
  → **New physics particles preferentially couple to the Higgs boson**
- Extended Higgs sector (MSSM, NMSSM etc.) allows the SM Higgs boson to act as a portal to a "hidden sector" of new physics interactions
- Run 2 focused on measuring the Higgs properties, including probes to BSM physics [3, 4]

New exotic phase space to be explored with additional data from Run 3

- Various SM Higgs couplings have only been constrained→new physics couplings may still be present
- Direct search for exotic particles is able to probe several TeV energy scales

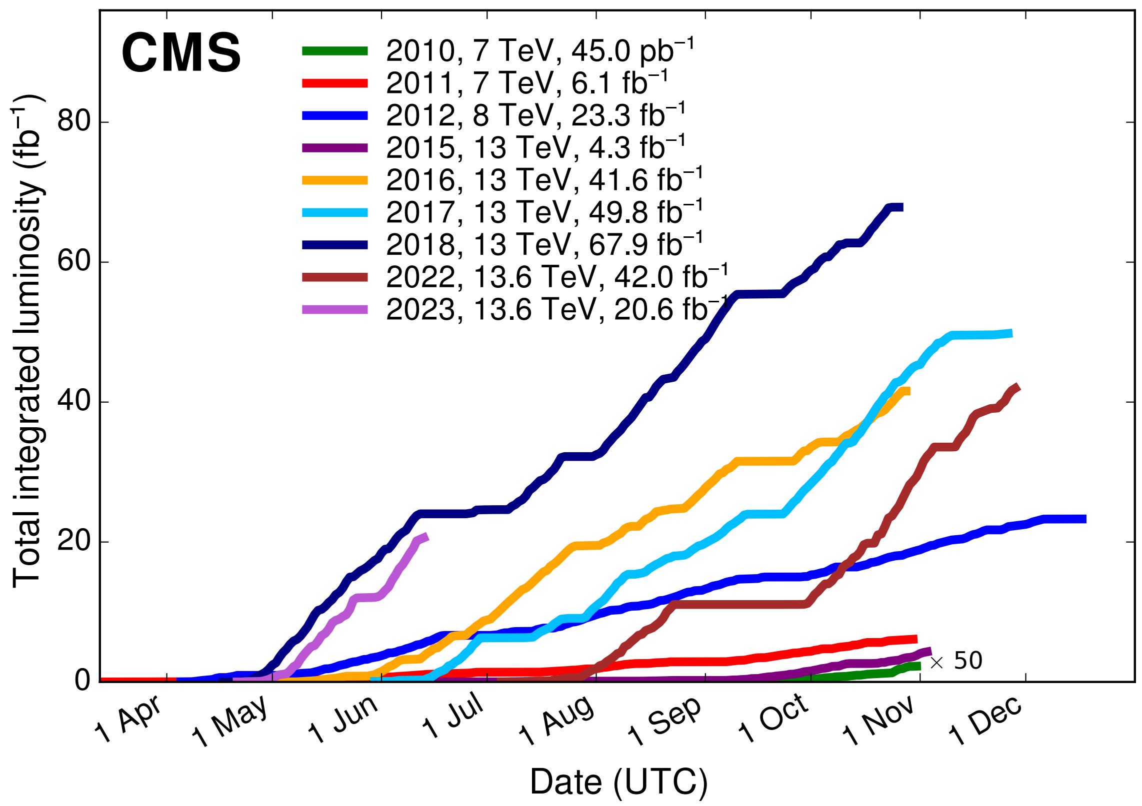


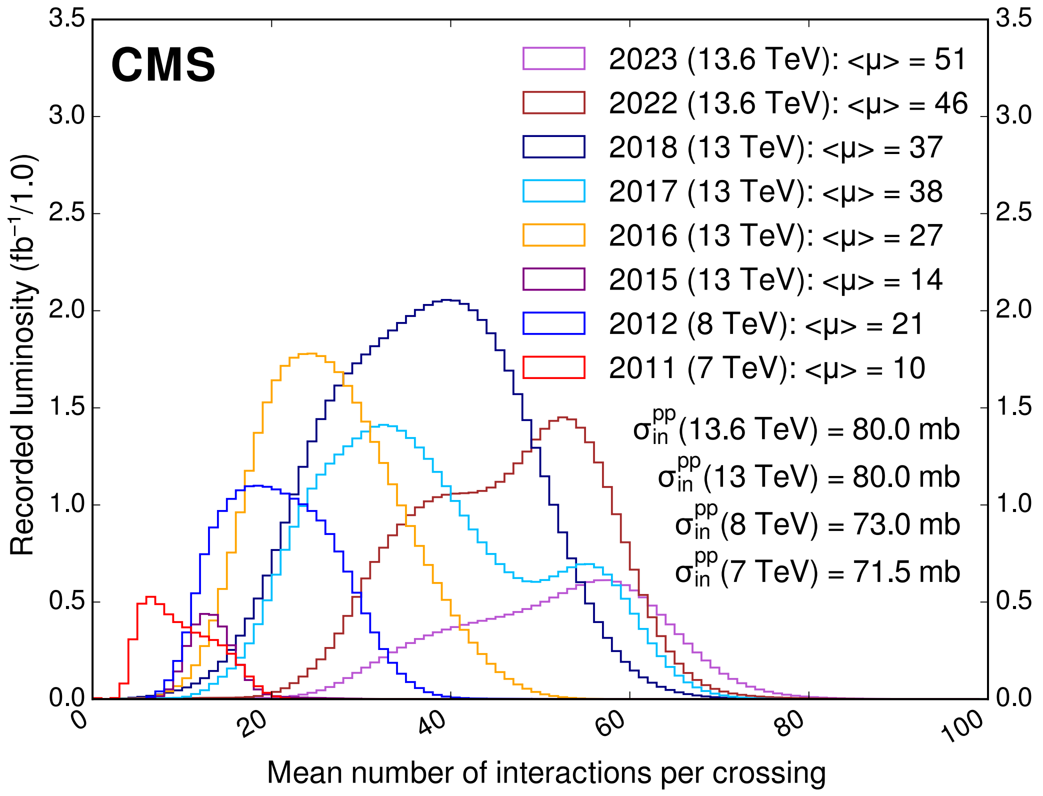


This talk: reviewing full Run-2 results of H→aa→2b2τ search from CMS-PAS-HIG-22-007

The Large Hadron Collider

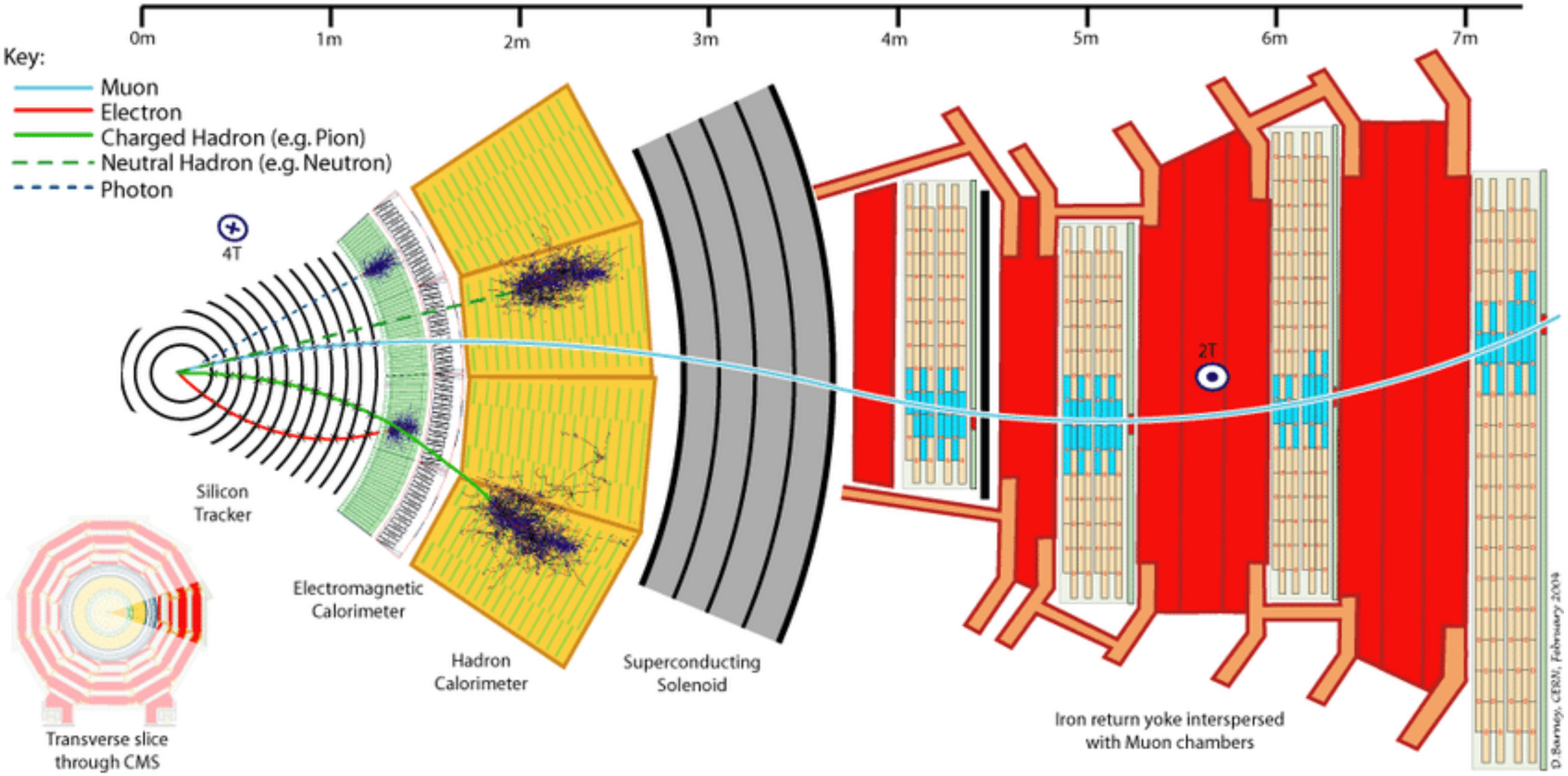
- World’s largest and most powerful particle collider in discovery mode
- Run 2 (2015-18): beam energy = 6.5 TeV and peak luminosity up to $1.5 * 10^{34}$ cm^{-2}s^{-1}
- Main physics goals:
  - Discover the Higgs boson and measure its properties
  - Search for beyond SM phenomena at TeV energy scale
- LHC is currently in Run-3 (2022-2024): beam energy = 6.8 TeV
- In future LHC will operate in the High Luminosity (HL) mode with a luminosity of about $7.5 * 10^{34}$ cm^{-2}s^{-1} and accumulate 3 ab^{-1} of collision data





Compact Muon Solenoid experiment

- One of the two general purpose detectors at the LHC, built around a superconducting solenoid
- Dedicated sub-detectors: silicon tracker, electromagnetic and hadronic calorimeters and muon system to identify and measure different particles
- Interesting physics events are selected online in two steps: Level-1 Trigger (hardware based) and High Level Trigger (computing farm)
- Combined information to reconstruct collision event → Particle Flow (PF)



Looking for a signal in collision events

- Low level reconstruction: hit positions or energy deposits in sub-detectors are combined to form track segments or energy clusters
- These are interpreted by PF algorithm as particle signatures: electrons, muons, photons, taus, jets, missing transverse momentum (p_T^{miss})
- Apply requirements to select events having expected signal-like features
  - This analysis: use Deep Neural Network to discriminate signal and background processes
- Extract total cross section by measuring observed data events:

$$\sigma^*BR(obs) = (N^{data} - N^{background}) / (\mathcal{L}^*A^*\epsilon)$$

- Compare σ*BR(obs) with σ*BR(theoretical) using the signal strength parameter:

$$\mu = \sigma^*BR(obs) / \sigma^*BR(theoretical)$$

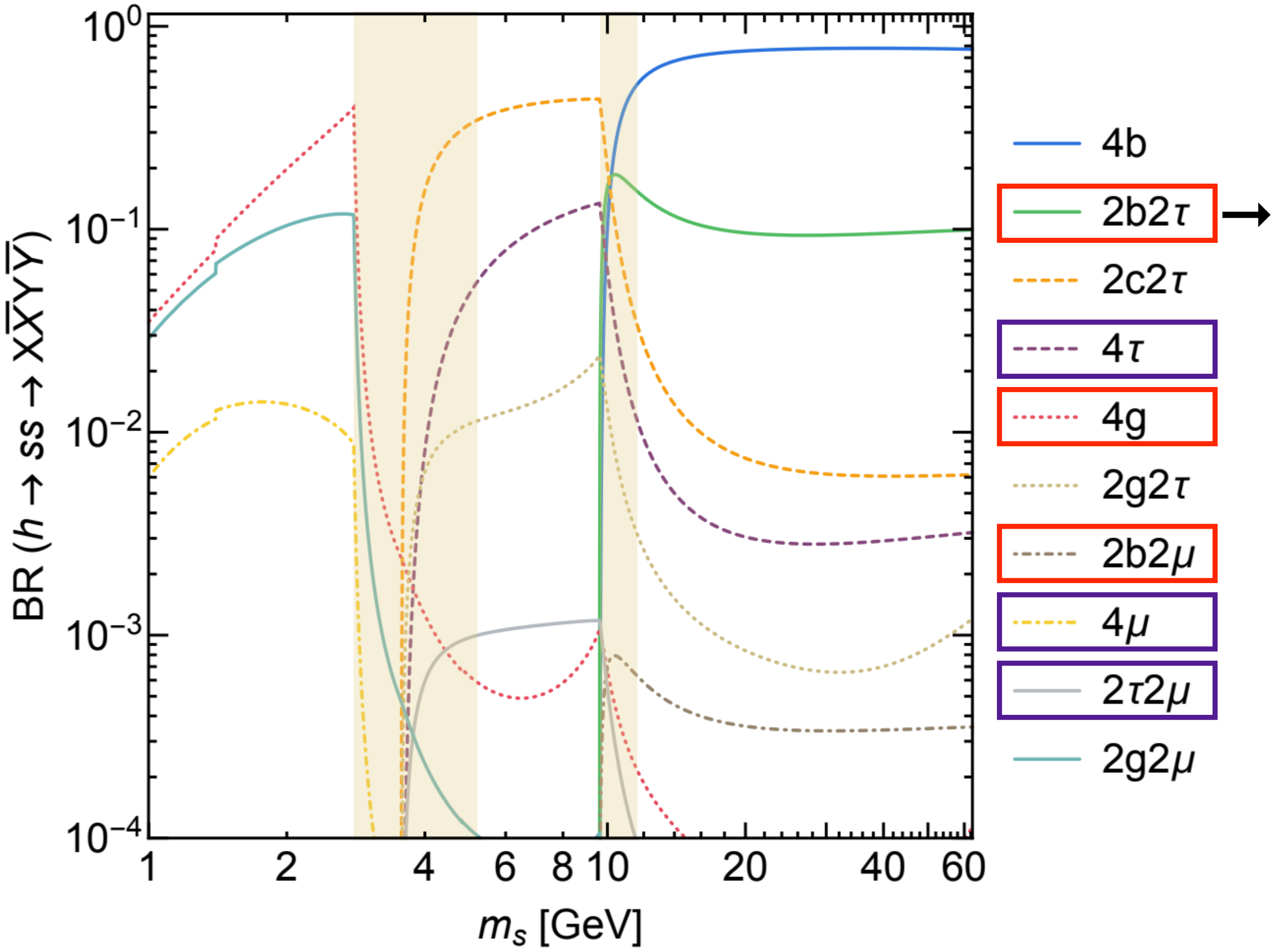
- Upper limit on μ is obtained using Maximum Likelihood fit approach, taking into account related experimental and theoretical uncertainties
  - This analysis: results are interpreted in terms of upper limits on BR($H \to aa \to 2\tau 2b$) by assuming SM Higgs production cross-section, and BR($H \to aa \to 2\tau 2b$) = 100%

Higgs to pseudoscalar decays

- Viable decay in 2HDM+S: two scalar doublets and one scalar singlet, leading to seven scalars or pseudoscalars
- Assuming the singlet state has no direct Yukawa couplings, decays to fermions are a result of mixing with the Higgs sector
- Mixing is small enough to preserve the SM couplings of the Higgs, branching fractions of the pseudoscalars depend on the model and model parameters
  → Different BSM models can be tested considering H→aa but special interest is in constraining 2HDM+S that conserve observed features of the SM

Predicted decay branching ratios of H to a decoupled singlet state (s) in 2HDM+S assuming BR(H→aa) is 100%

arxiv:1312.4992



→ this talk

Full Run-2 results

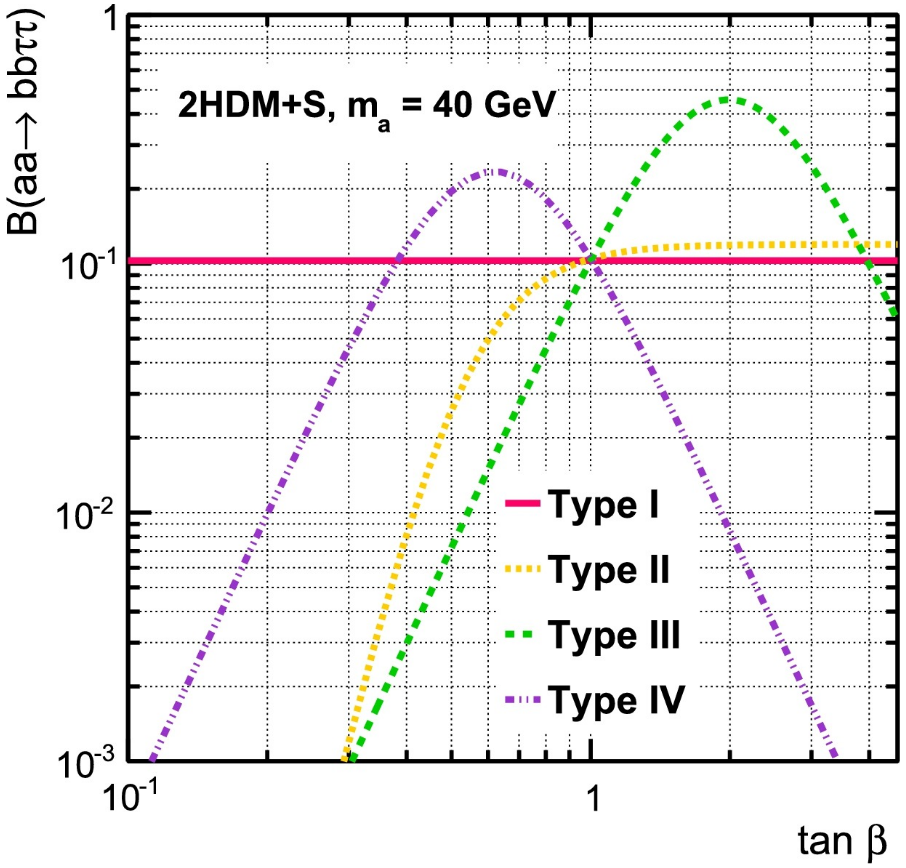
Partial Run-2 results

2HDM+S

- Four types of 2HDM+S are defined which forbid FCNC, based on coupling structure of the two Higgs doublets and the SM fermions

	Type I	Type II	Type III (lepton specific)	Type IV (flipped)
Right handed leptons	h_1	h_2	h_2	h_1
Up-type quarks	h_1	h_1	h_1	h_1
Down-type quarks	h_1	h_2	h_1	h_2

- Different BR is predicted depending on the model type and tanβ value (ratio of vacuum expectation values)
  - Highest production rate of $H \rightarrow aa \rightarrow 2\tau 2b$ is predicted by the Type III model
  - Type II scenario most interesting in terms of phenomenology: default coupling structure for most MSSM theories



ANALYSIS

H→aa→2τ2b analysis in a nutshell

Relatively larger BR to bb and ττ, improved τ lepton reconstruction techniques

- Search for a masses within 12 $< m_a <$ 60
- Three final states explored: $e\mu$, $e\tau_h$, $\mu\tau_h$

Improved results compared to the previous analysis using partial Run-2 data (2016)

- Addition of > 1 b-jet category made possible due to increased statistics
- DNN categorisation vs. cut based event selection strategy
- SVfit algorithm to reconstruct di-tau invariant mass $m_{\tau\tau}$ including neutrino energies instead of only visible components of $m_{\tau\tau}$ distribution
- Better object reconstruction techniques based on DNN developed within CMS experiment in the recent years: DeepJet, DeepTau tagging
- More precise estimation of Z→ττ using the embedding technique

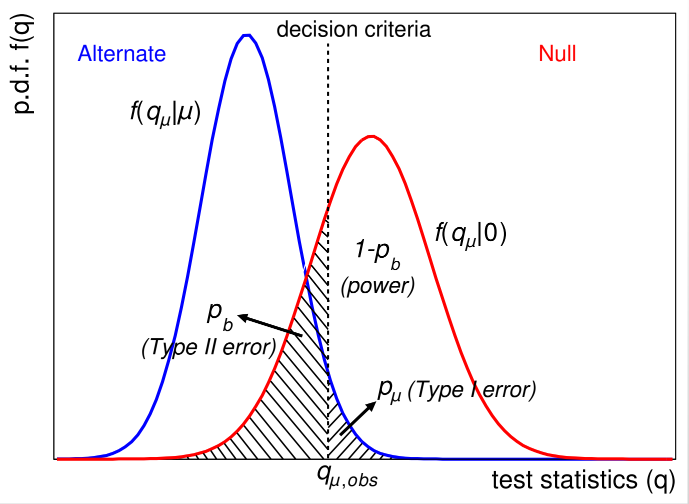
2016-only result: BR(H→aa→2τ2b) values constrained at 95% CL below 3-12% depending on m_a

The expected upper limits from the full Run-2 analysis improved due to changes in analysis strategy rather than the increase in data statistics alone

Analysis workflow

Data/Sim. → Trigger for each of eμ, eτ_h, $\mu\tau_h$ → Object selection (e, μ, τ_h, b-jet) → Event selection based on b-jet multiplicity → Six DNN trainings → Event categorisation based on DNN score → Shape analysis including systematics (Maximum Likelihood fit)

Consider different theoretical models → Shape analysis including systematics (Maximum Likelihood fit)



Shape analysis including systematics (Maximum Likelihood fit)

Shape analysis → Limits on BR(H→aa→2τ2b)

Shape analysis → Post-fit distributions

Trigger requirements and object selection

		$e\mu$		$e\tau_h$		$\mu\tau_h$	
	Type	e	μ	e	τ_h	μ	τ_h
2016	single	-	-	25	-	22	-
	leading	23	23	-	-	-	20
	sub-leading	12	8	-	-	19	-
2017	single	-	-	27, 32	-	24, 27	-
	leading	23	23	-	30	-	27
	sub-leading	12	8	24	-	20	-
2018	single	-	-	32, 35	-	24, 27	-
	leading	23	23	-	30	-	27
	sub-leading	12	8	24	-	20	-

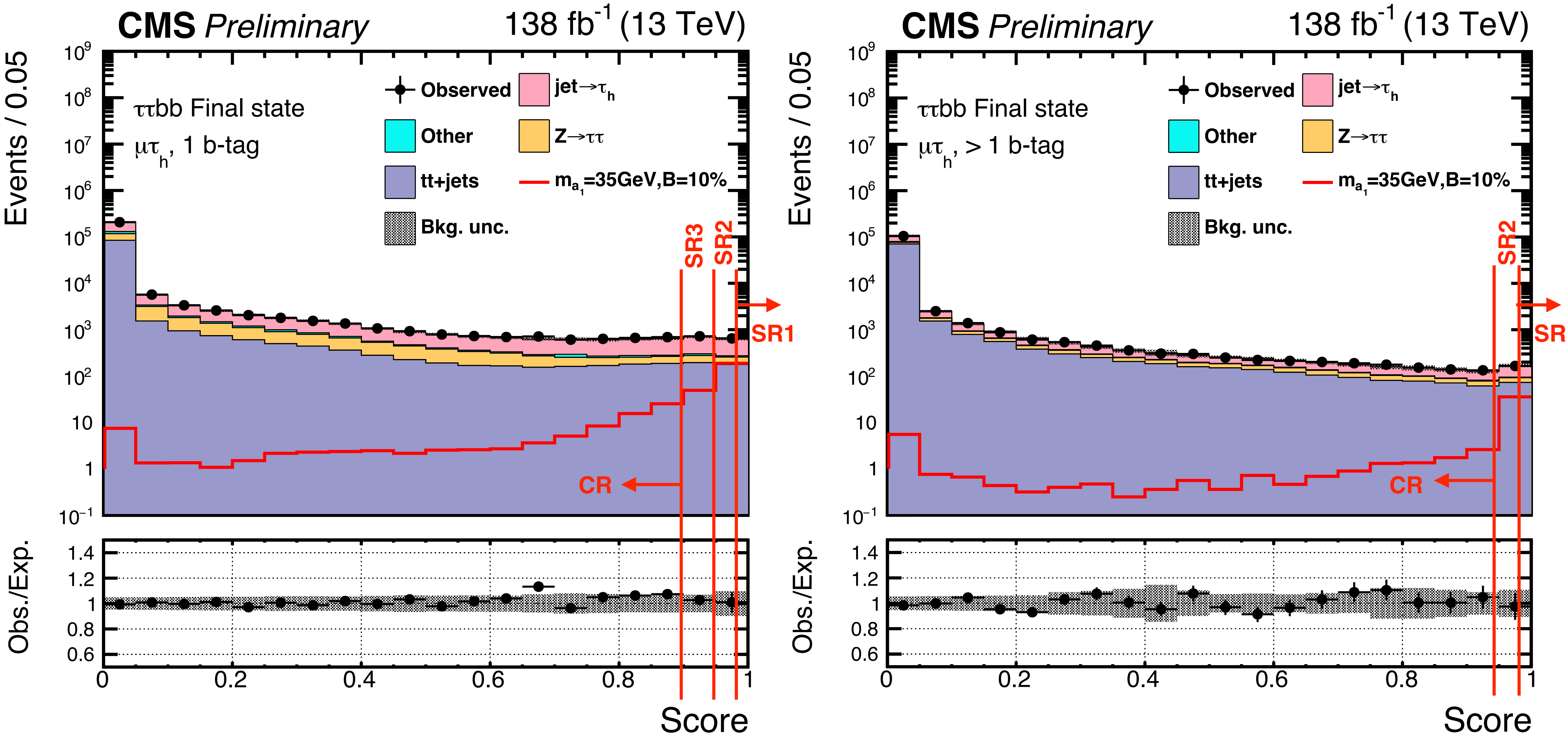
- Electrons and muons are reconstructed within $|\eta| < 2.4$ and τ_h within $|\eta| < 2.1$
- Offline e, μ and τ_h are matched to the trigger objects, with p_T thresholds being 1 GeV larger than the online threshold for e, μ; offline p_T threshold for τ_h is 35 GeV
- In case both single and cross-triggers are present in the event, use lowest threshold
- Additional identification/isolation requirements on e/μ/τ_h (e.g. DeepTau for taus)
- Anti-kT jets are reconstructed within $|\eta| < 2.4$ using a cone size of 0.4; they are required to have $p_T > 20$ GeV; b-tagged using DeepJet algorithm

Event selection

- Only three di-tau final states considered:
  - ee and $\mu\mu$ have low BR and large background from Drell-Yan process
  - $\tau_h\tau_h$ has high trigger threshold
  - Extra lepton veto applied for each of the three final states to ensure mutually exclusive selection
- Events should have at least one loosely tagged b-jet with $p_T > 20$ GeV
  - Two broad categories based on b-jet multiplicity: = 1 and > 1 b-jet
- DNN categorisation:
  - Discriminate signal against a combination of major backgrounds ($t\bar{t}$+jets and Drell Yan)
  - Train one DNN for each of the three channels and two b-jet categories: six in total
  - Training variables are based on kinematics of reconstructed final state particles
  - Split the selected events further into smaller categories based on the DNN scores: in some of these categories more signal events are selected compared to the background enhancing sensitivity called the signal regions (SR); background rich categories are taken as control regions (CR)

DNN score distributions

- Discriminating observables include invariant mass of visible decay products, transverse mass between an object and p_T^{miss}, m_{bb}-$m_{\tau\tau}$ etc.
- Final observable used in maximum likelihood fit: $m_{\tau\tau}$, not used as an input to DNN



Background estimation

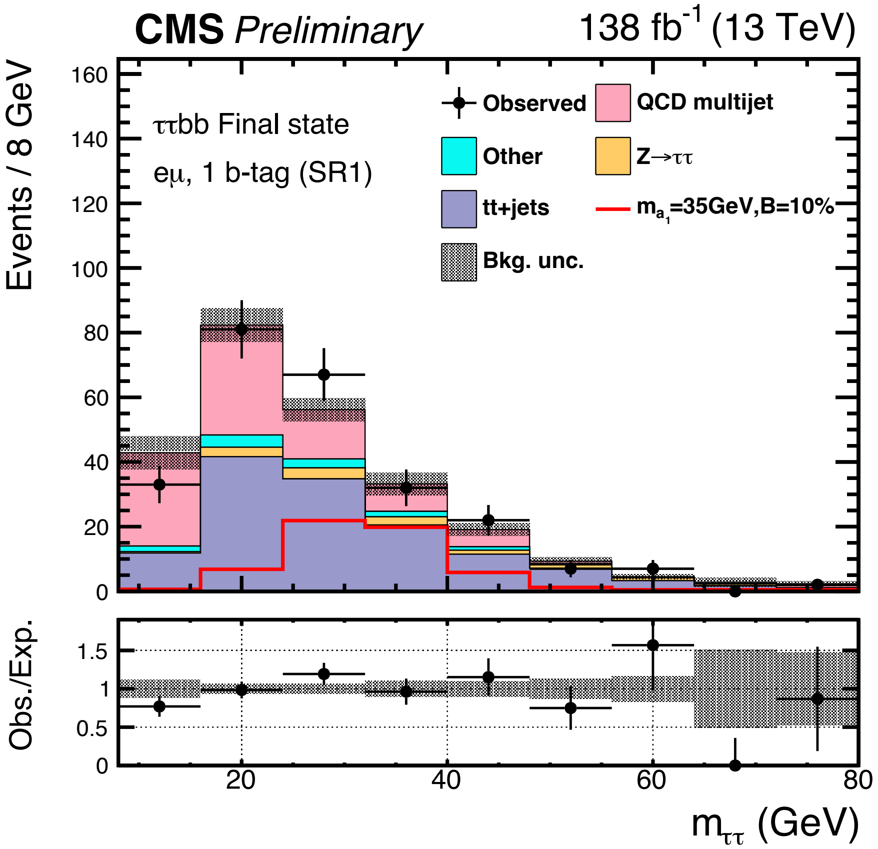
- Irreducible physics backgrounds: genuine particles forming the final state from other physics processes are estimated from simulation
  - Z→ee/$\mu\mu$
  - W+jets in the eμ channel
  - $t\bar{t}$+jets in the eμ channel
  - Diboson, single top, SM Higgs→ττ/WW
- Reducible backgrounds: mis-identified or *fake* particles forming the final state are estimated from data, also Z→ττ that is not described well in simulation
  - Jets faking τ_h: W+jets and QCD processes have large jet multiplicity, leading to fake τ_h; they are estimated from a sideband region in data by multiplying with a scale factor derived from control region; this estimate also includes contributions from $t\bar{t}$+jets, single top, diboson and Z+jets
  - QCD process in eμ channel: jets can also be mis-identified as e/μ and are most significant in QCD process, which has high jet multiplicity; thus it is estimated in a similar way using scale factors derived from control region
  - Z→ττ: the limitations in reconstructing taus is overcome by the embedding technique; well reconstructed Z→$\mu\mu$ events are selected from data and the muon candidates are replaced with simulated tau candidates having the same kinematics → includes better description of jets and detector conditions

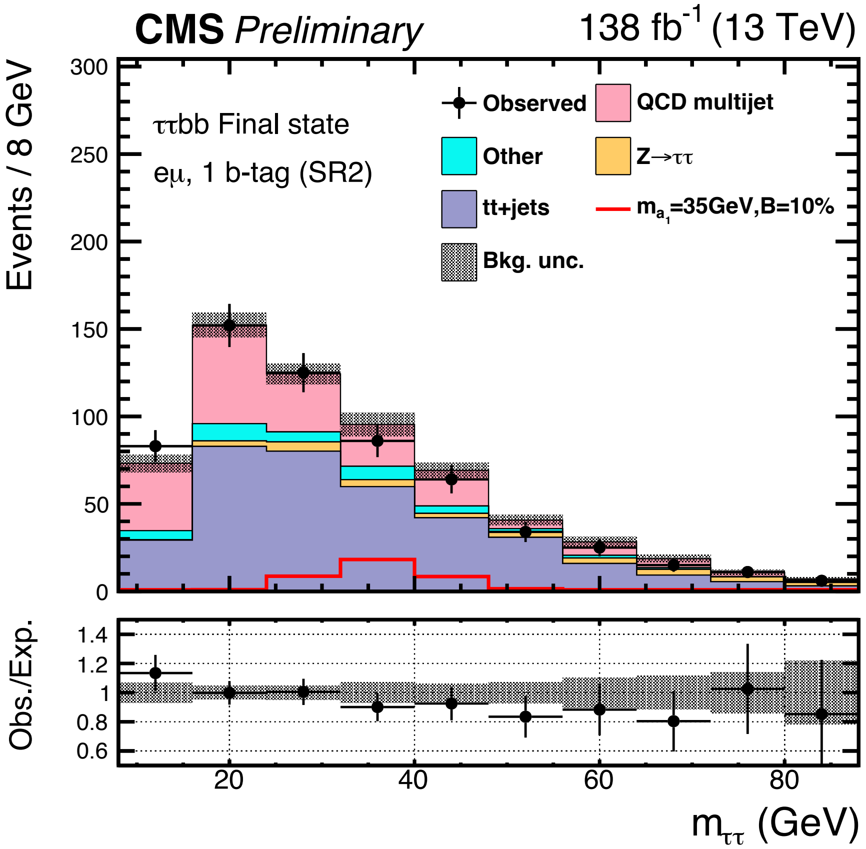
Systematic Uncertainties

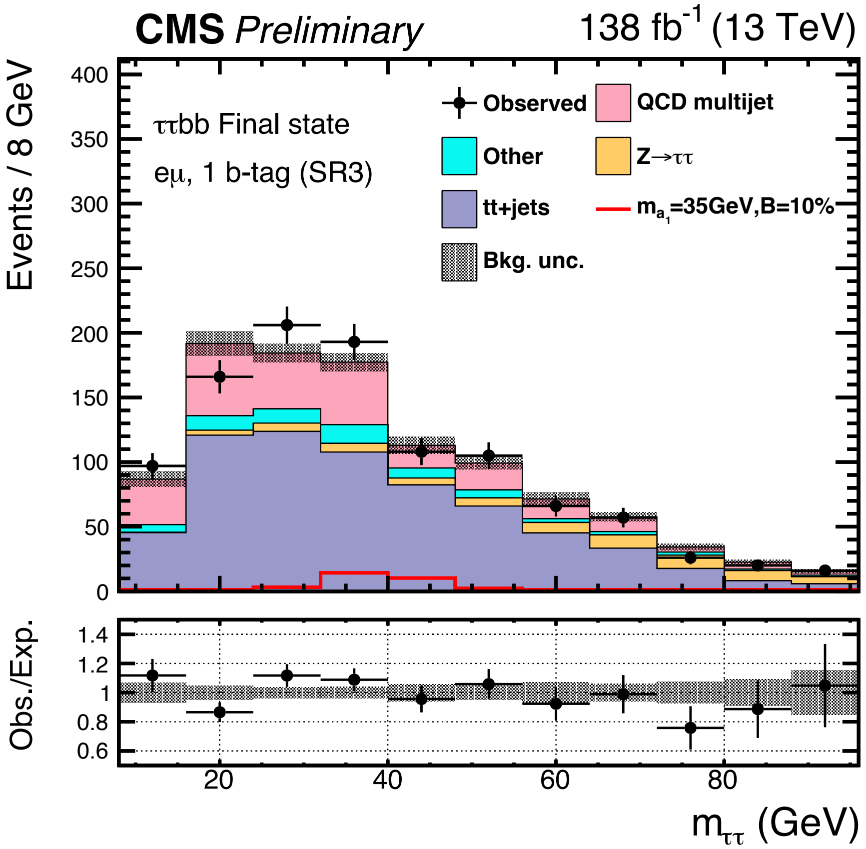
- Some uncertainties affect the shape of the $m_{\tau\tau}$ distribution, some only vary the yield
- Two broad categories: experimental and theoretical
- Experimental:
  - Luminosity measurement
  - Uncertainty in measuring efficiency scale factors for $e/\mu/\tau_h$ selection and trigger
  - Jet energy correction and b-tagging efficiencies
  - ECAL timing shift due to misalignment
  - Background estimations:
    - Normalisation of various SM process
    - Uncertainty in measuring different fake rates/scale factors for data-driven backgrounds
    - Uncertainty in estimating the embedded background
- Theoretical:
  - Uncertainty in the ggF and VBF production cross sections of the Higgs boson
  - Scale variations in $t\bar{t}$+jets, single top and diboson simulations
  - Parton-shower uncertainties in $t\bar{t}$+jets

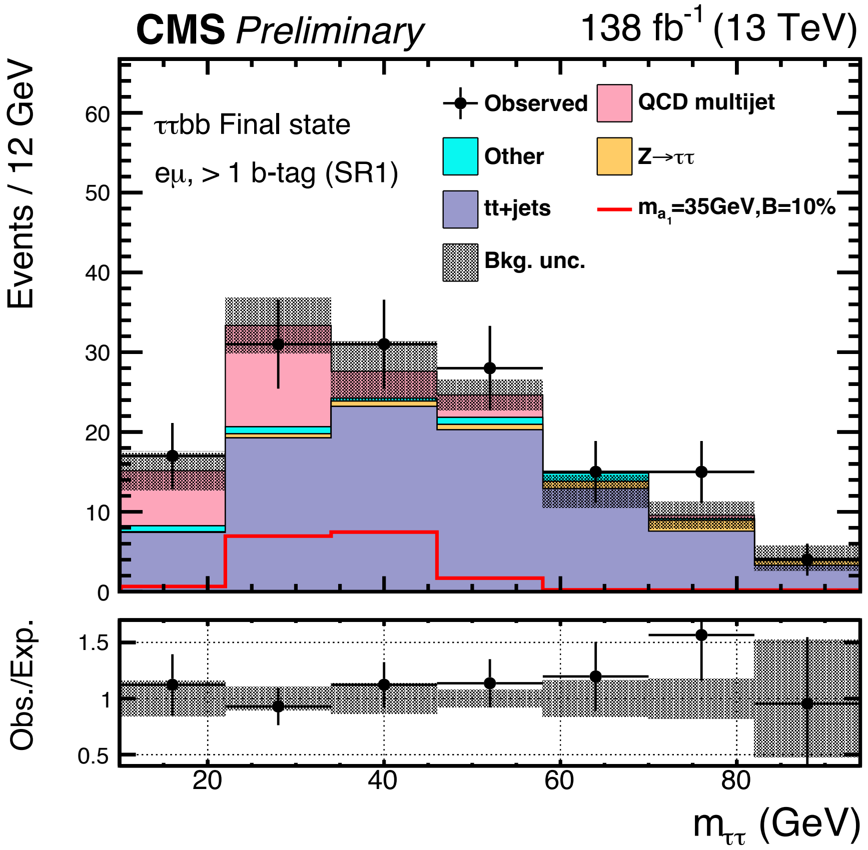
RESULTS

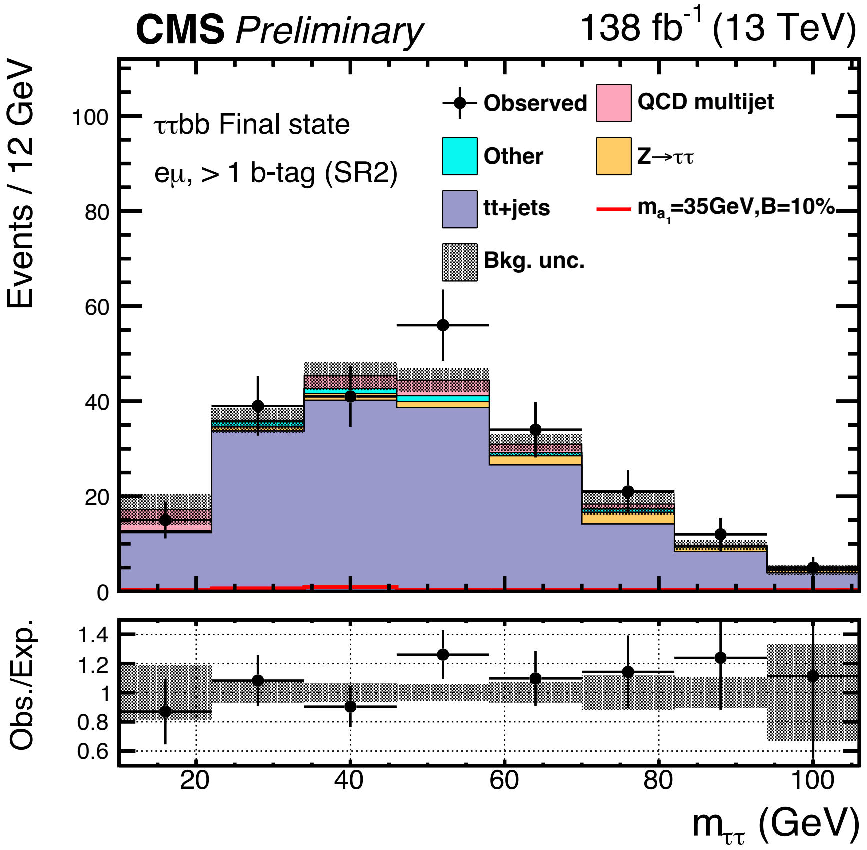
Single region distributions: eμ



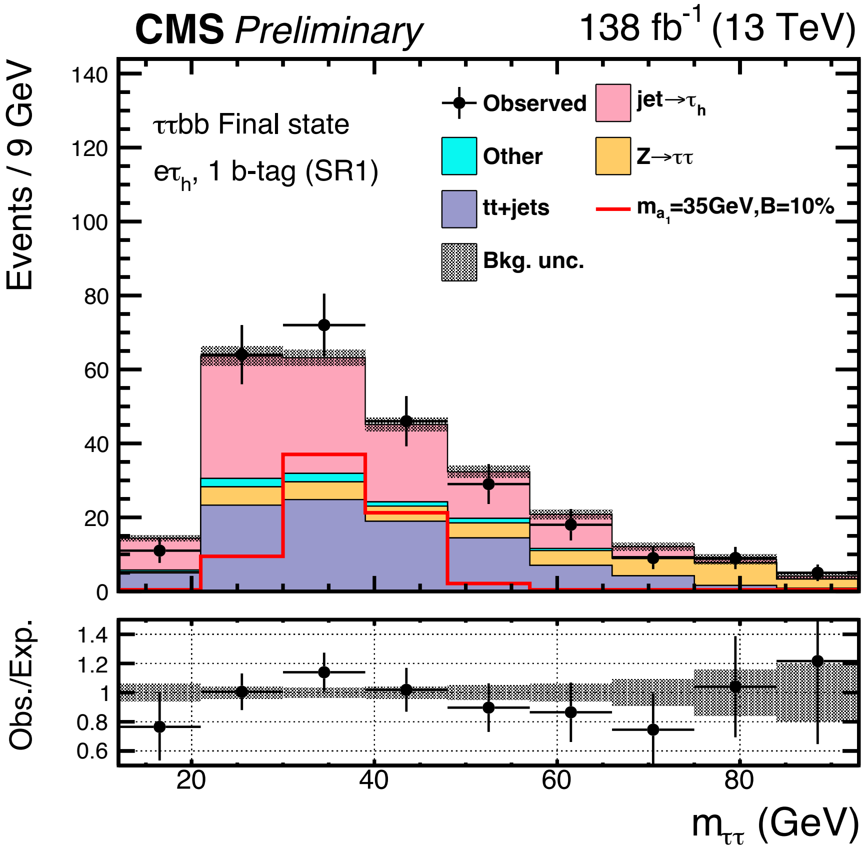


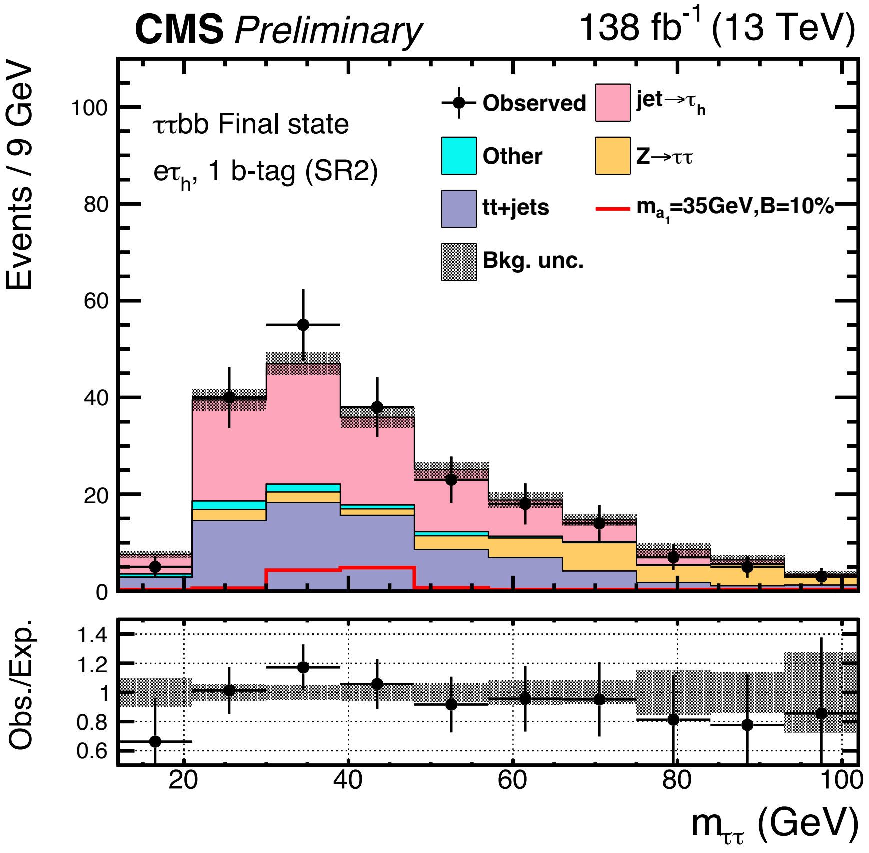


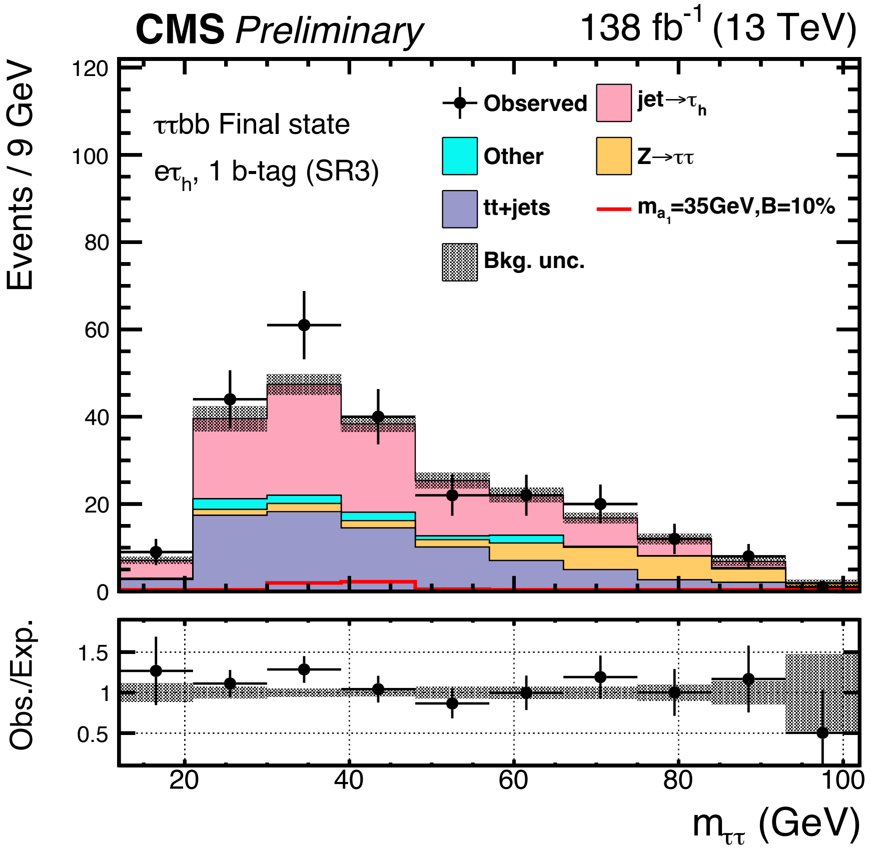


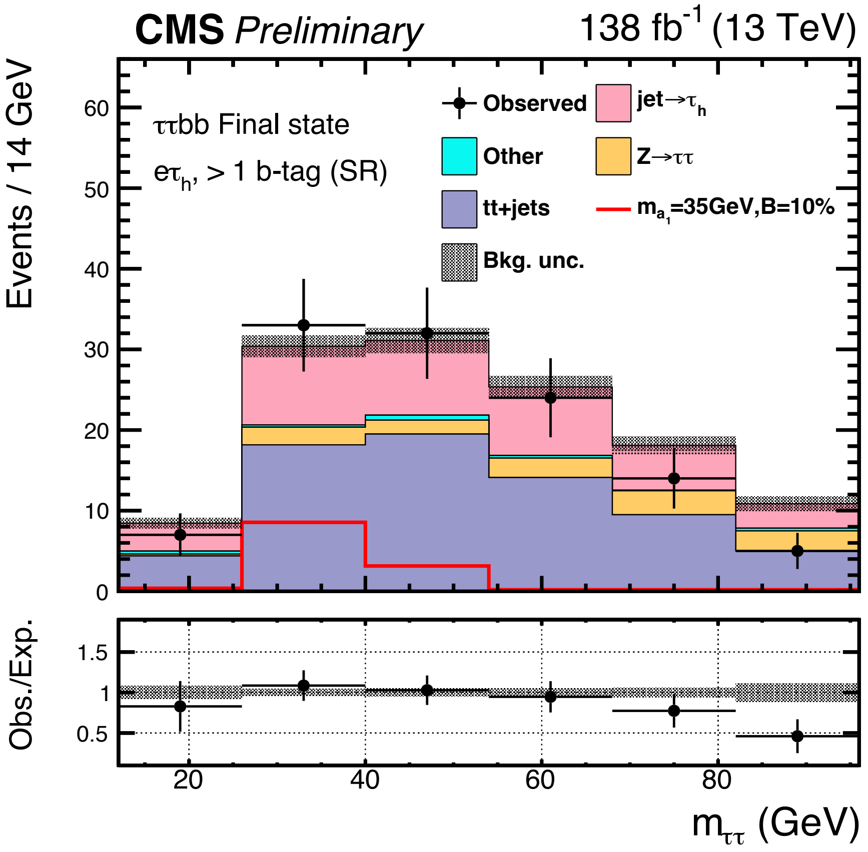


Single region distributions: $e\tau_h$

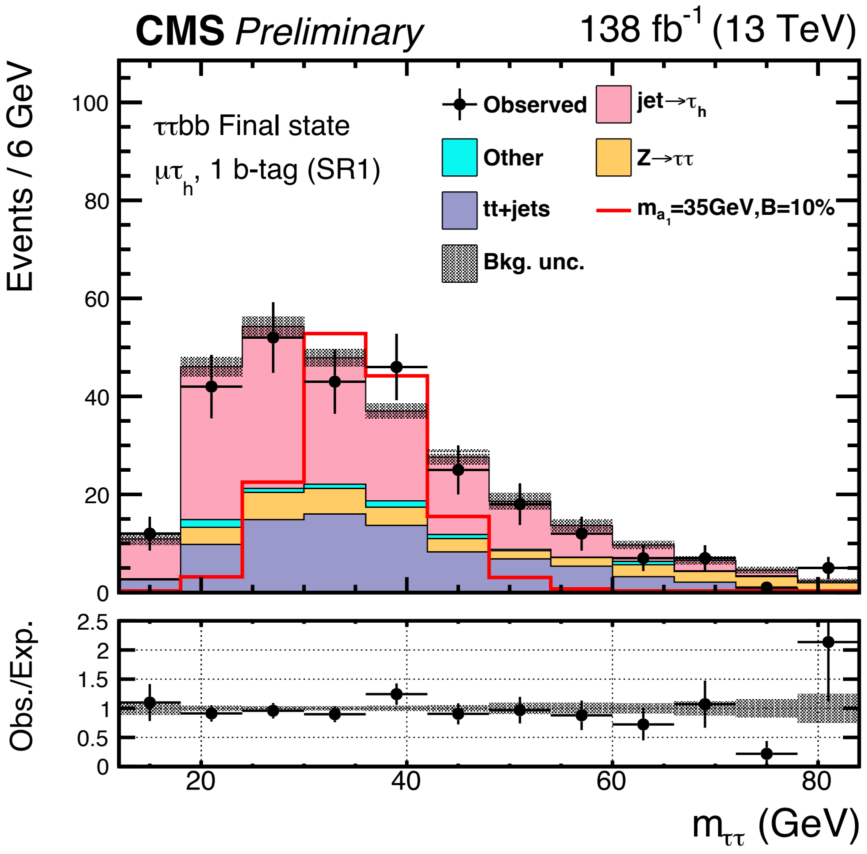


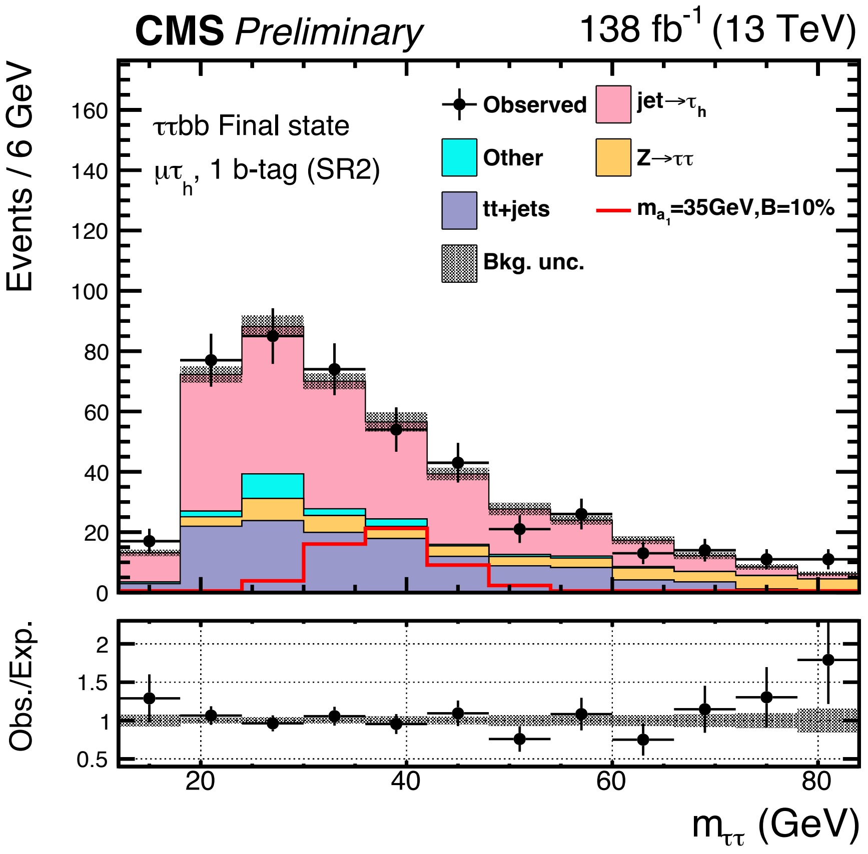


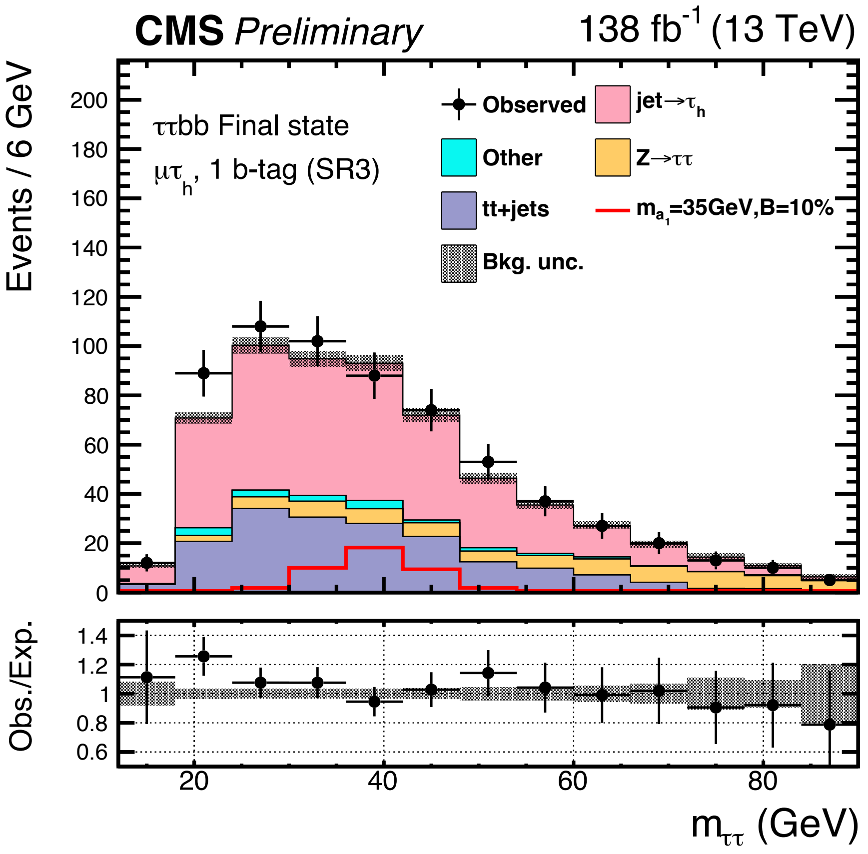


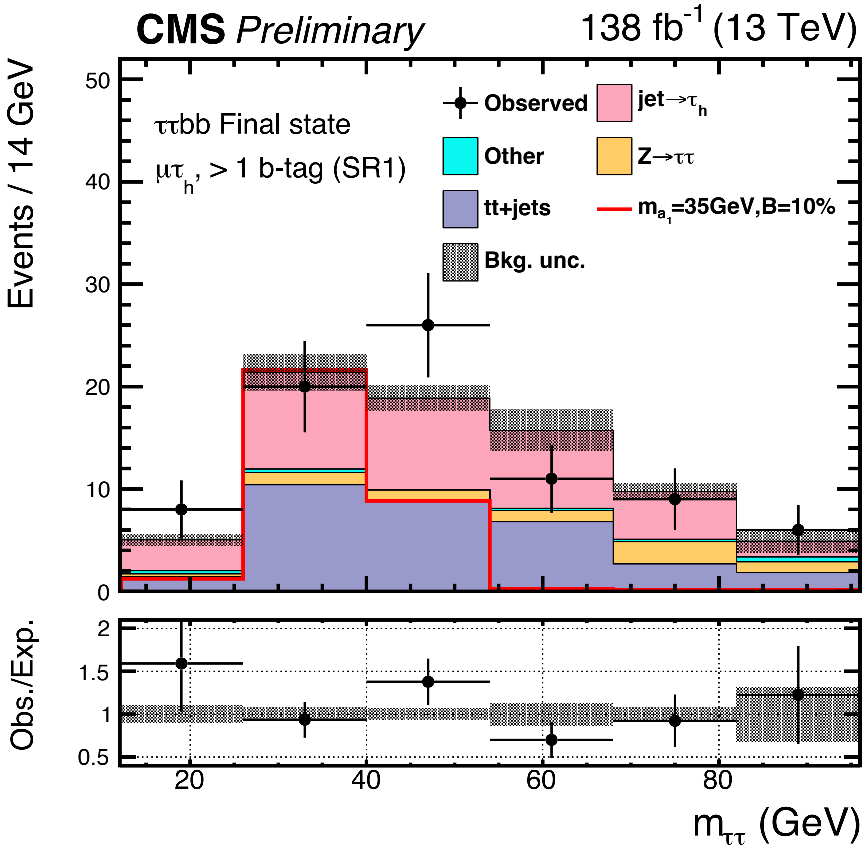


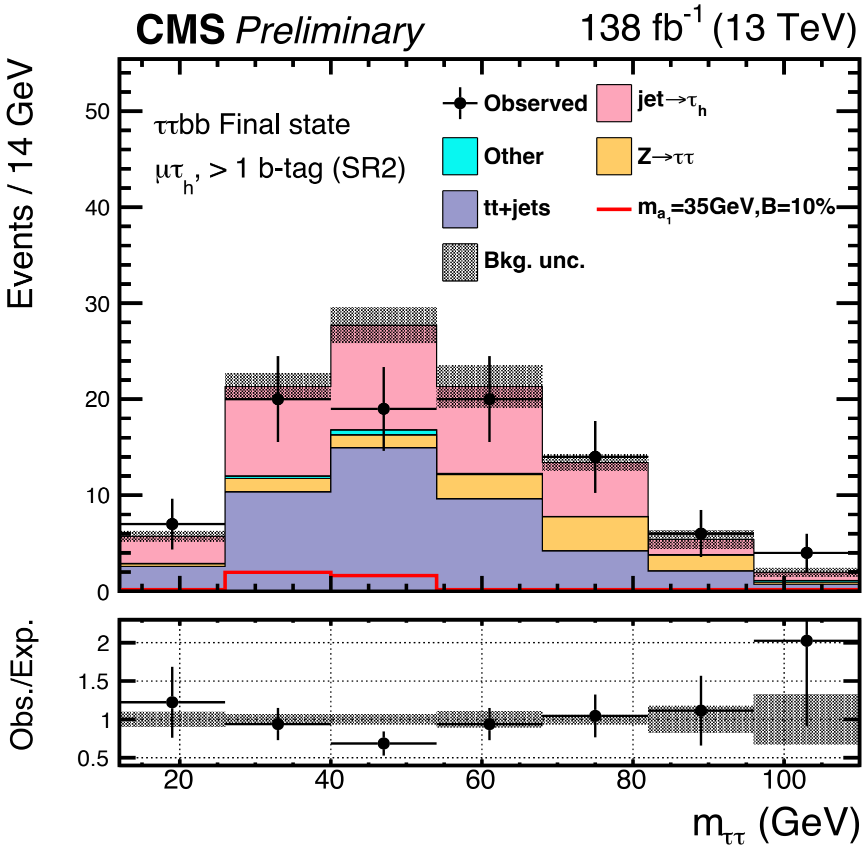
Single region distributions: $\mu\tau_h$







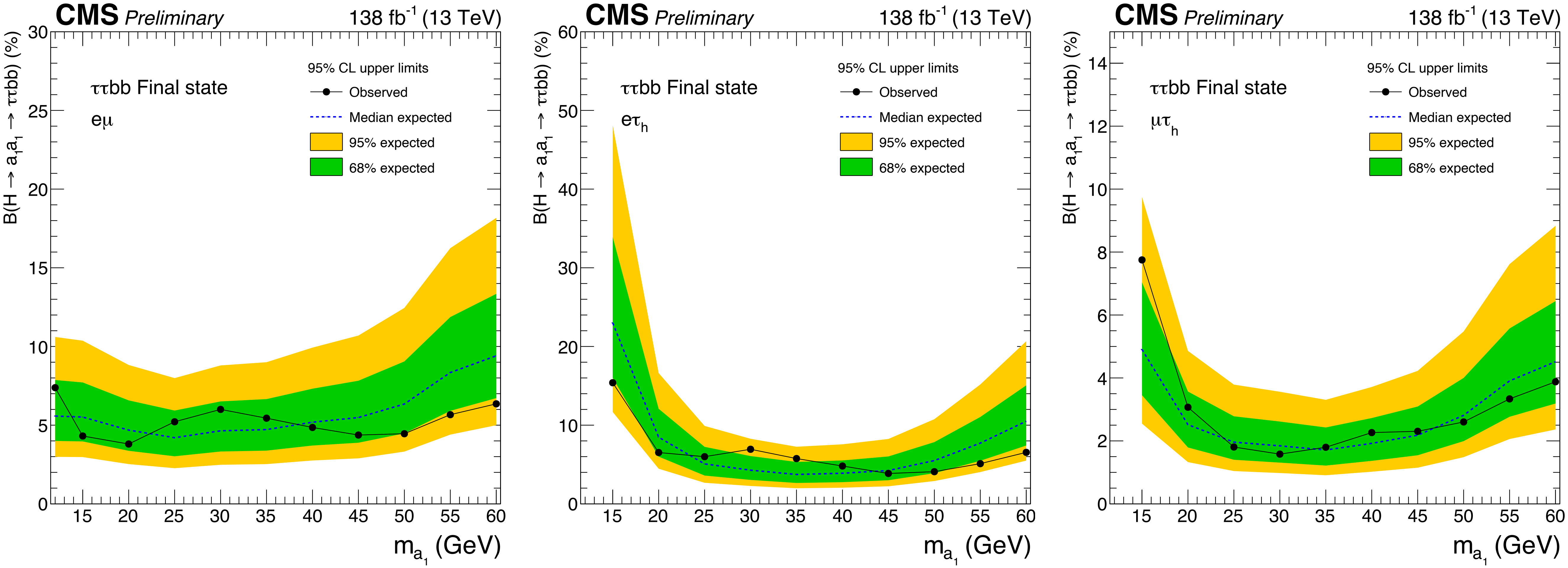




Upper limits on exotic Higgs BR

Limit is set on SM like Higgs→aa→2τ2b:

- Most sensitive channel: $\mu\tau_h$, dominant background is Z→ττ and τ_h fakes from QCD multijet
- Dominant systematic uncertainty from fake τ_h background estimation
- Analysis is still statistically limited

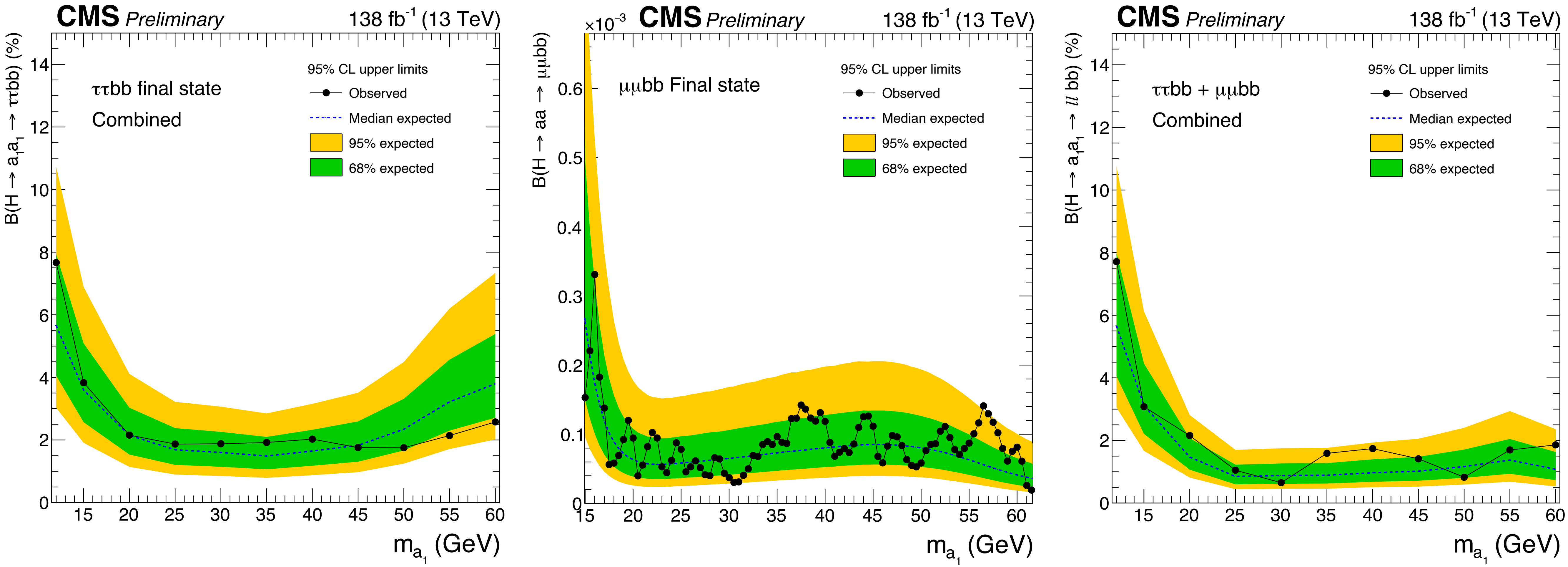


Only the $e\mu$ channel is sensitive to the 12 GeV mass point

- For low m_a the decay products are boosted, need dedicated reconstruction
- In this analysis, a ΔR requirement is applied between the final state particles, which has a lower threshold in $e\mu$ channel

Combination with H→aa→2μ2b analysis

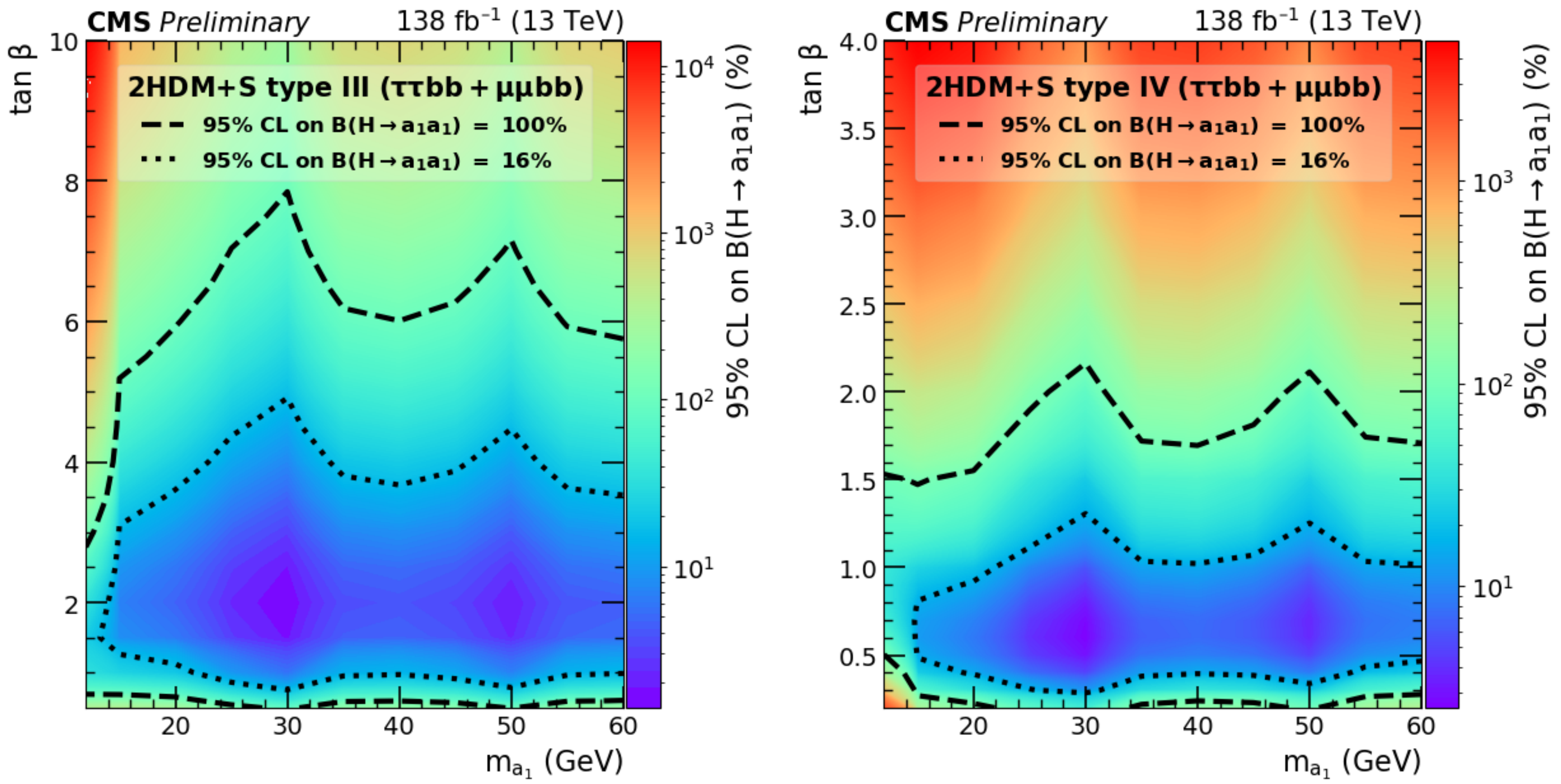
- Straightforward statistical combination: analyses utilise orthogonal data samples
- Some common uncertainties are treated as correlated, such as luminosity measurement, jet energy scale, variations in signal cross section etc
- Type-independent upper limits on BR(H→aa→llbb) in the context of 2HDM+S are derived as a function of m_a where l is a μ or τ



Interpreting in terms of different 2HDM+S: BR(H→aa) values excluded above 23% (Type II tanβ > 1), 7% (Type III tanβ = 2.0) and 15% (Type IV tanβ = 0.5)

Implications for different models

Stringent upper limits are set for most Type III and Type IV 2HDM+S scenarios



16% contour corresponds to combined upper limit on Higgs to BSM particle decays obtained from previous Run 2 results

TL; DL…

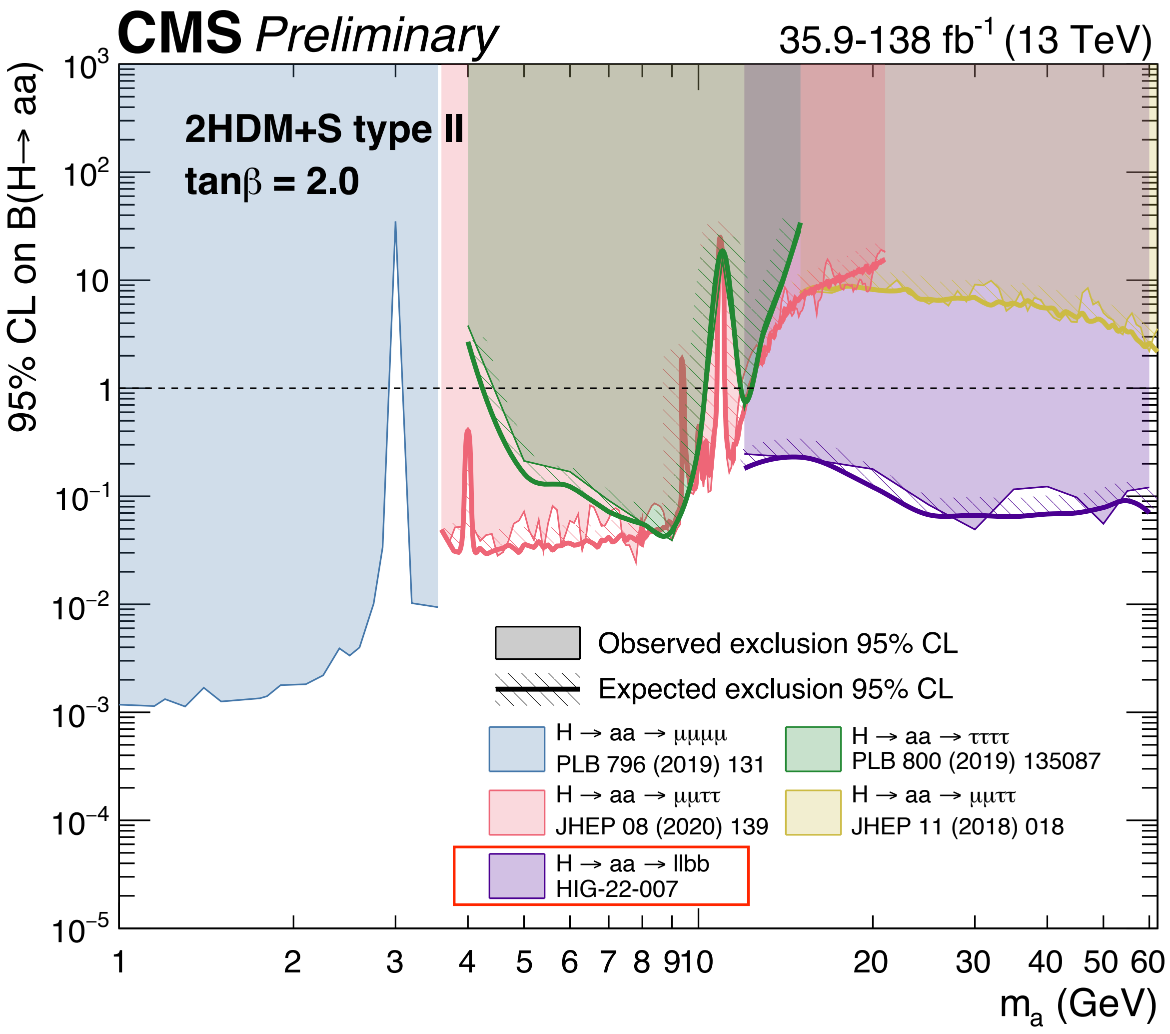
Summary

Higgs portal to hidden BSM sector being explored by CMS analyses in different final states

→ **Many full Run-2 results are public, some are work in progress**

- Improved sensitivity compared to previous searches using novel analysis techniques and machine learning
- For H→aa→2τ2b, no significant excess over SM prediction *just yet*, many other possibilities remain to be explored
  - Asymmetric pseudoscalar masses unexplored
  - Signals with low pseudoscalar mass to be analysed using boosted reconstruction techniques

Direct searches benefit the most with increase in luminosity: exciting times ahead with the onset of LHC Run-3!



Thank You

Backup

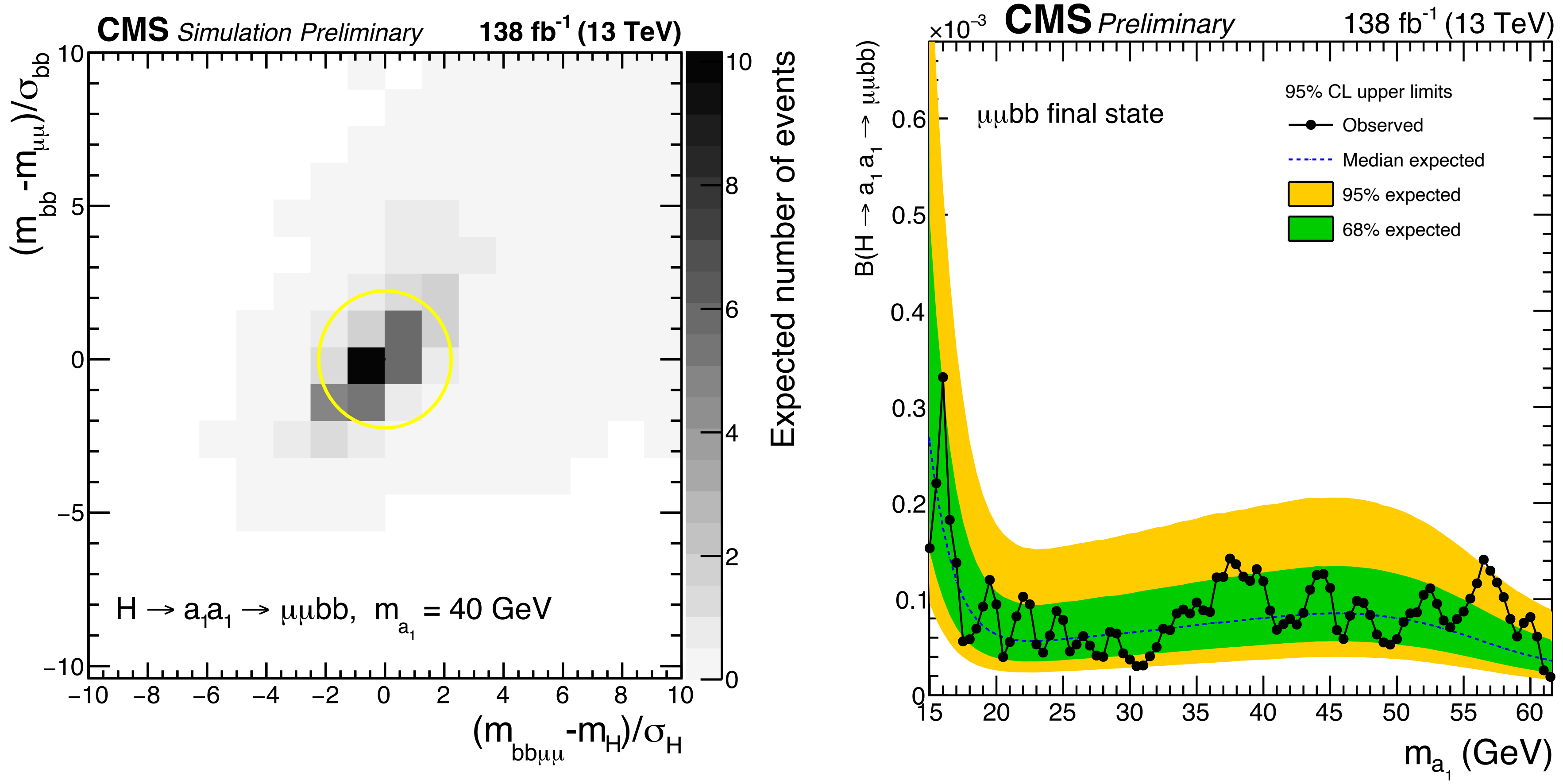
$H \rightarrow aa \rightarrow 2\mu 2b$

Clean signature with a precise mass resolution from $m_{\mu\mu}$ and large BR from bb

- Search for a masses within 15 < m_a < 60
- Bump hunt analysis using the dimuon invariant mass $m_{\mu\mu}$
- Completely data-driven background estimation
- Thorough study of the signal to use a single discriminating variable to suppress background

Parametric fit of the signal model in different categories based on b-jet properties

Most stringent observed upper limit till date in this final state, slightly better than ATLAS results

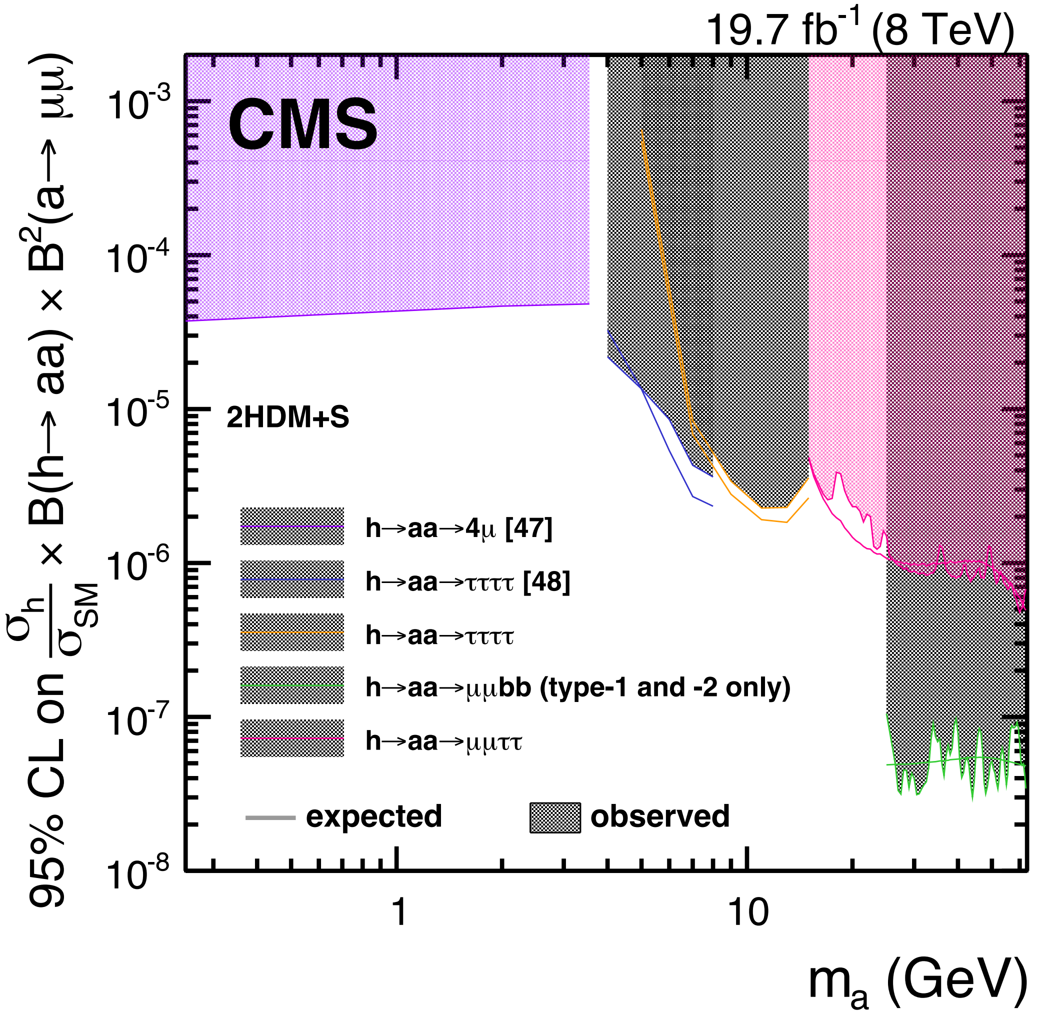


No significant deviations from SM prediction, analysis is limited by statistics

Run-1 results: $h \to aa \to l^+l^-, l^+l^-$

Results from Run-1: using 19.7 fb^{-1} p-p collision data at 8 TeV

- Pseudoscalar masses between 5 and 62.5 GeV are probed in final states 4τ, $2b2\mu$, and $2b2\tau$
- Results were compared to predictions from 2HDM and 2HDM+S models



$B(a \to \tau^+\tau^-)$ is directly proportional to $B(a \to \mu^+\mu^-)$ in any type of 2HDM+S, as is $B(a \to bb)$ in Type-1 & -2

Therefore, the results of all analyses can be expressed as exclusion limits on $\sigma(h)/\sigma$ SM x $B(H \to aa)B^2(a \to \mu^+\mu^-)$

SM compatibility: combining ATLAS and CMS measurements an upper limit of 34% is set on exotic Higgs decays → loose constraint on BSM physics