TOP2024 Workshop
Saint-Malo

Search for heavy scalar or pseudoscalar states in $t\bar{t}$ events at CMS

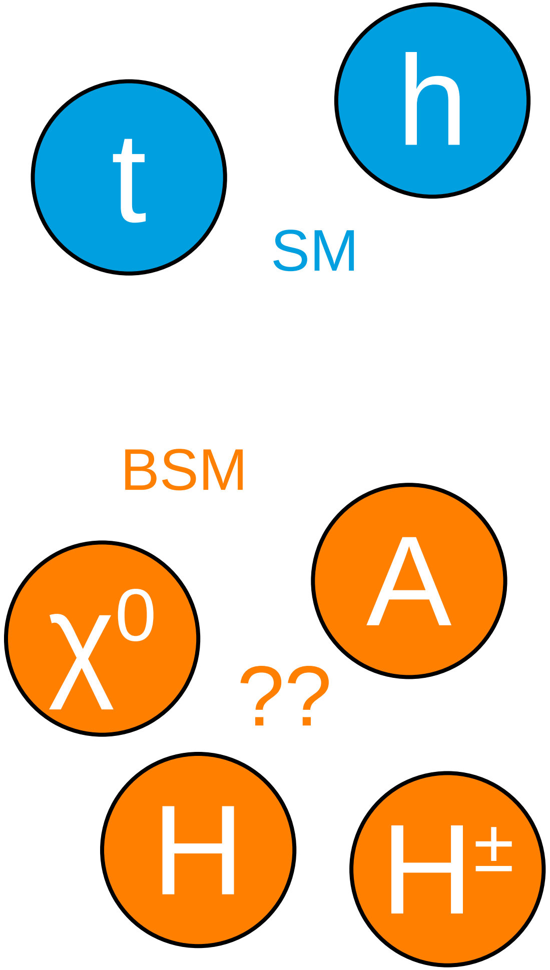
Laurids Jeppe for the CMS collaboration

26.09.2024 | CMS-PAS-HIG-22-013

Motivation: BSM!

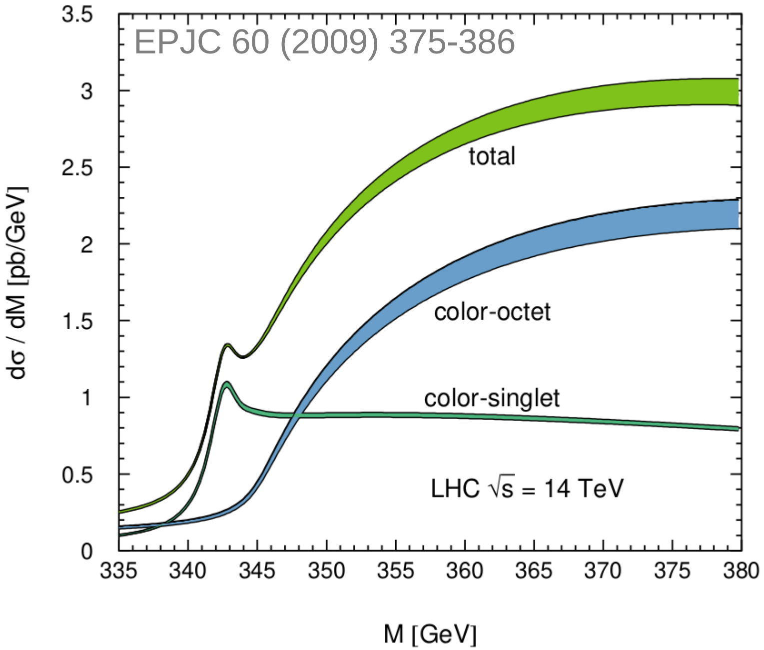
- We know there is BSM – but where is it?
- Many models predict extended Higgs sectors e.g. 2HDM, MSSM…
  $\rightarrow$ new scalar or pseudoscalar states
- If couplings are Yukawa-like: strongest coupling to top quark
- If mass > $2m_t$: decay to $t\bar{t}$ is dominant in large areas of parameter space

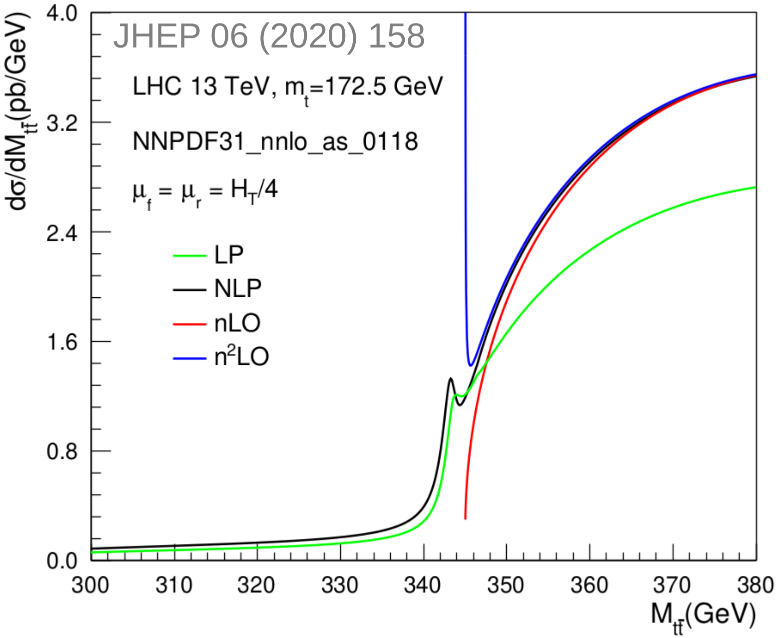
$\rightarrow$ search for heavy (pseudo)scalars in $t\bar{t}$ final states



Motivation: $t\bar{t}$ bound states?

- SM predicts $t\bar{t}$ (quasi-)bound states below the $t\bar{t}$ threshold



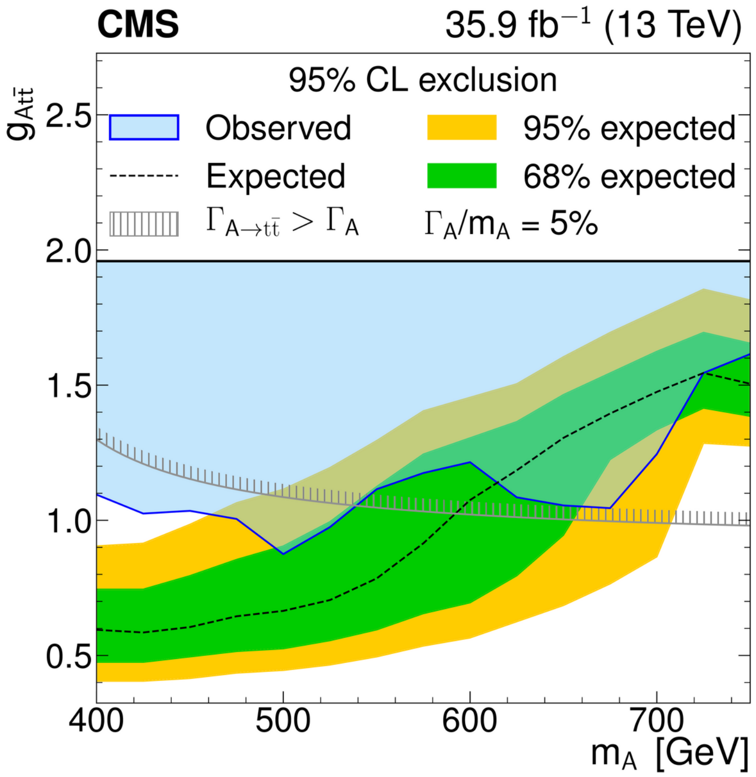


- Not observed yet (but there are hints: differential $t\bar{t}$, entanglement...)
- Dominant component: pseudoscalar – can we search for it?

Overview of the search

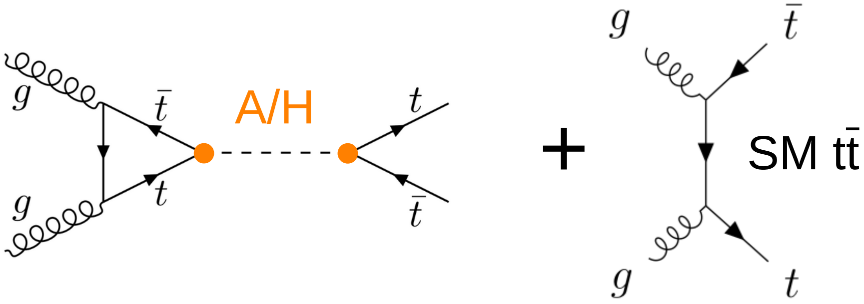
- Search for new spin-0 (pseudo)scalars in $t\bar{t}$ final states with full Run 2 dataset (138 fb^{-1})
- Make use of invariant $t\bar{t}$ mass, angular and spin correlation observables
- Two analysis channels: dilepton (ℓℓ) and lepton+jets (ℓj)
- Builds upon previous work by CMS: JHEP 04 (2020) 171 (2016 data, 35.9 fb^{-1})
- Also recent full Run 2 result by ATLAS (JHEP08 (2024) 013) - Eleanor's talk on Tuesday

JHEP 04 (2020) 171



Signal modeling

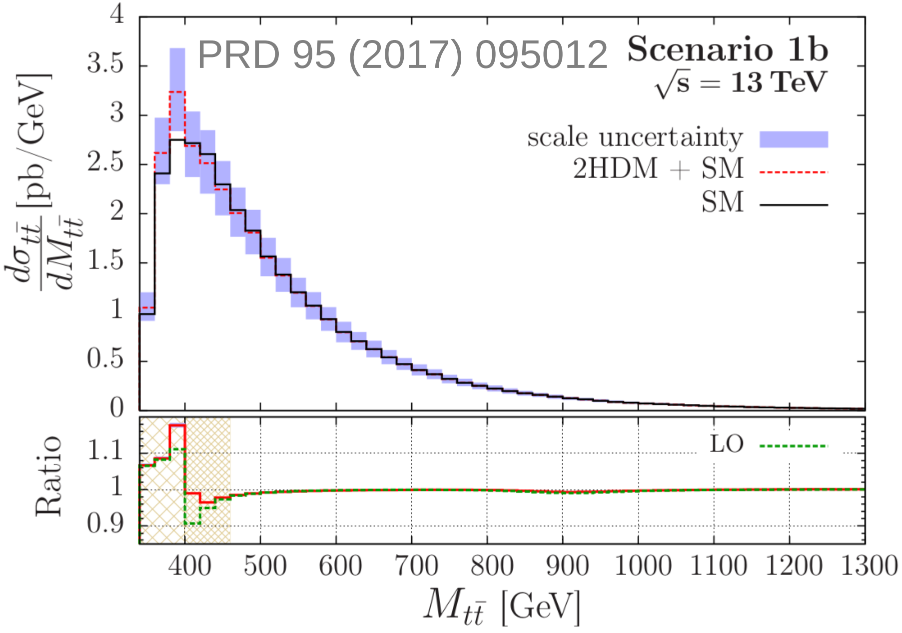
- Generic heavy pseudoscalar (A) or scalar (H) coupling solely to top quarks
- Production in gluon fusion via top quark loop



- Same final state as SM $t\bar{t}$ $\rightarrow$ interference $\rightarrow$ peak-dip structure in m_{tt}
- Free parameters: masses, widths, coupling modifiers g_A / g_H
- Use NNLO QCD K-factors for normalization

$$\mathcal{L}_A^{\text{int}} = i g_{\text{At}\bar{\text{t}}} \frac{m_{\text{t}}}{v} \bar{\text{t}} \gamma_5 \text{t} \text{A}$$

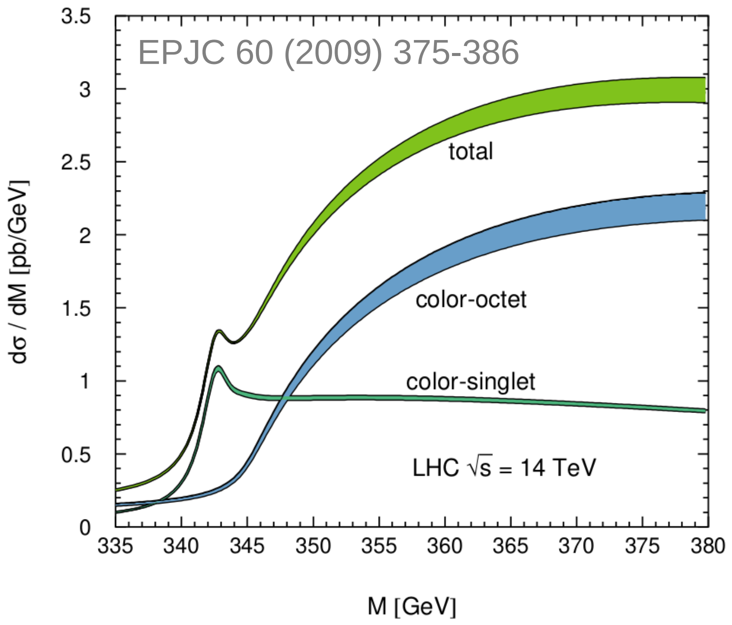
$$\mathcal{L}_H^{\text{int}} = - g_{\text{Ht}\bar{\text{t}}} \frac{m_{\text{t}}}{v} \bar{\text{t}} \text{t} \text{H}$$



Modeling: $t\bar{t}$ bound state effects

See e.g.
PRD 110 (2024) 5, 054032
JHEP 03 (2024) 099
PRD 104 (2021) 3, 034023
etc.

- State of the art: non-relativistic QCD (NRQCD)
  - Color-singlet ($^1S_0^{[1]}$) - attractive
    $\rightarrow$ Peak below the $t\bar{t}$ threshold
  - Color-octet ($^1S_0^{[8]}$ or $^3S_1^{[8]}$) - repulsive
    $\rightarrow$ Expected to be small below the $t\bar{t}$ threshold



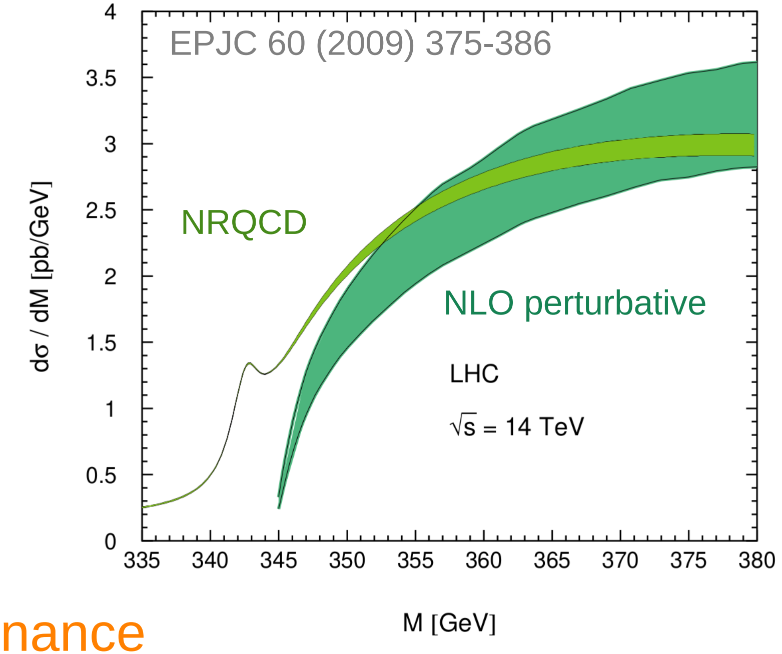
- Lineshape and width not exactly known
  - but below experimental resolution

Modeling: $t\bar{t}$ bound state effects

- Use simplified model for MC simulation: η_t (PRD 104 (2021) 034023)
  - Generic spin-0, color-singlet state η_t
  - Couplings to gluons and tops (pseudoscalar)
  - Fit mass from NRQCD:

$$m_{\eta_t} - 2m_t = -2\,\mathrm{GeV} \quad \Rightarrow \quad m_{\eta_t} = 343\,\mathrm{GeV}$$

  - Restrict to $m_{\mathrm{WbWb}} \in [337, 349]\,\mathrm{GeV}$ to not influence $t\bar{t}$ continuum as predicted by perturbative QCD
  - Not available yet: by-event reweighting to NRQCD



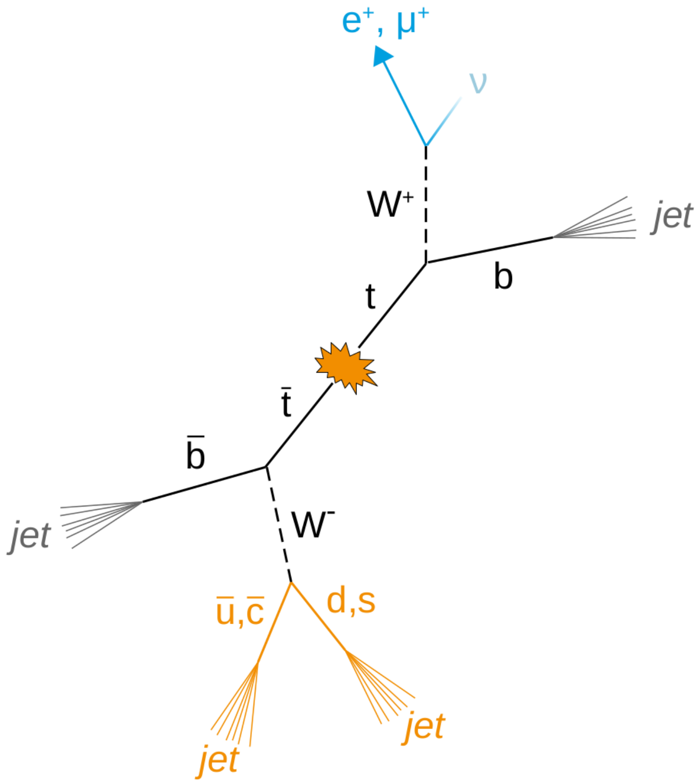
- Result: very similar signature as low-mass A resonance

Lepton+jets channel

- Require exactly one lepton (e/μ), 3 or more jets and 2 or more b tags
- Split into 4 categories: e vs μ and 3 jets vs 4+ jets

(NIM A 736 (2014) 169-178)

- Reconstruct $t\bar{t}$ system with NeutrinoSolver algorithm:
  - Assign b jets by maximum likelihood
  - Energy correction factor applied for 3 jet events (lost or merged jets) (NIM A 788 (2015) 128-136)

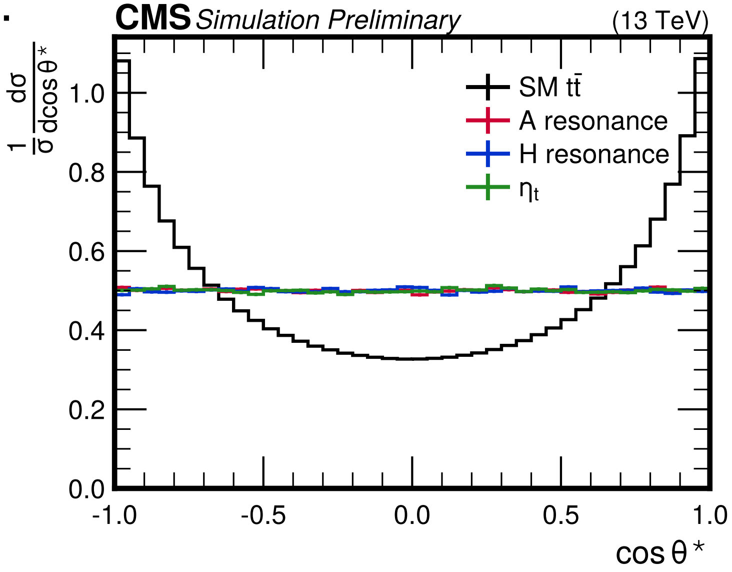


Lepton+jets channel

- Require exactly one lepton (e/μ), 3 or more jets and 2 or more b tags
- Split into 4 categories: e vs μ and 3 jets vs 4+ jets

(NIM A 736 (2014) 169-178)

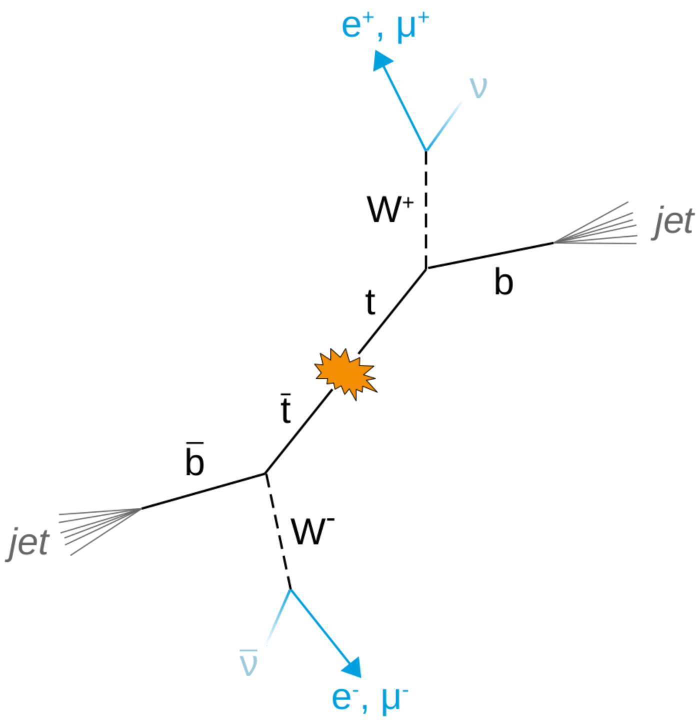
- Reconstruct $t\bar{t}$ system with NeutrinoSolver algorithm:
  - Assign b jets by maximum likelihood
  - Energy correction factor applied for 3 jet events (lost or merged jets) (NIM A 788 (2015) 128-136)
- 2D binning in m_{tt} x $|\cos\theta^*|$
- θ*: scattering angle of leptonic top quark
  - SM $t\bar{t}$: peaks at large angles
  - A/H signal: isotropic → flat distribution



Dilepton channel

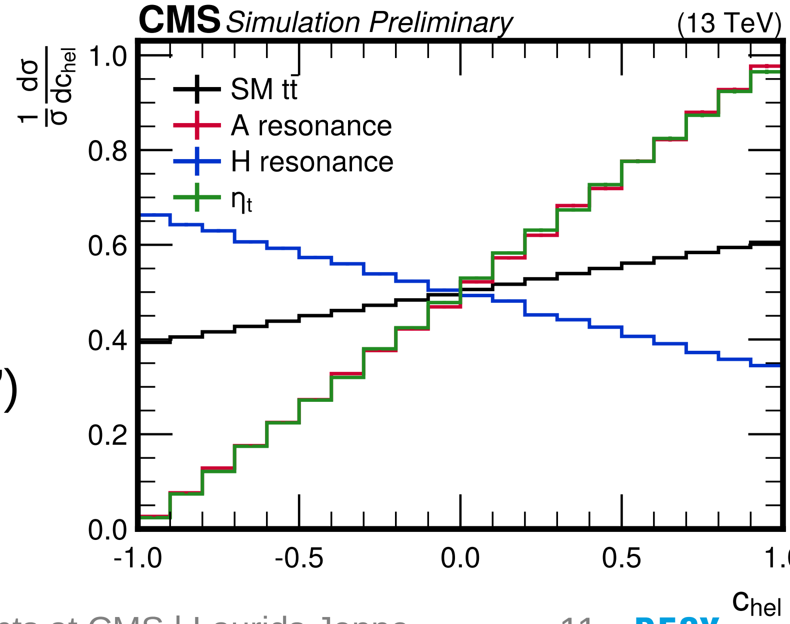
- Exactly two opposite-sign leptons (e/μ), at least 2 jets, and at least 1 b tag
- Split by lepton flavor: ee, eμ and μμ
- Reject low-$m_{\ell\ell}$ events
  Cut away Z peak & require p_T^{miss} > 40 GeV in ee/μμ

- Analytic reconstruction of $t\bar{t}$ system:
  - Assumptions: all p_T^{miss} from νν, tops and Ws on-shell
  - Assign b jets using likelihood based on $m_{\ell b}$
  - Finite detector resolution: repeat reconstruction 100 times with randomly smeared inputs, take weighted average



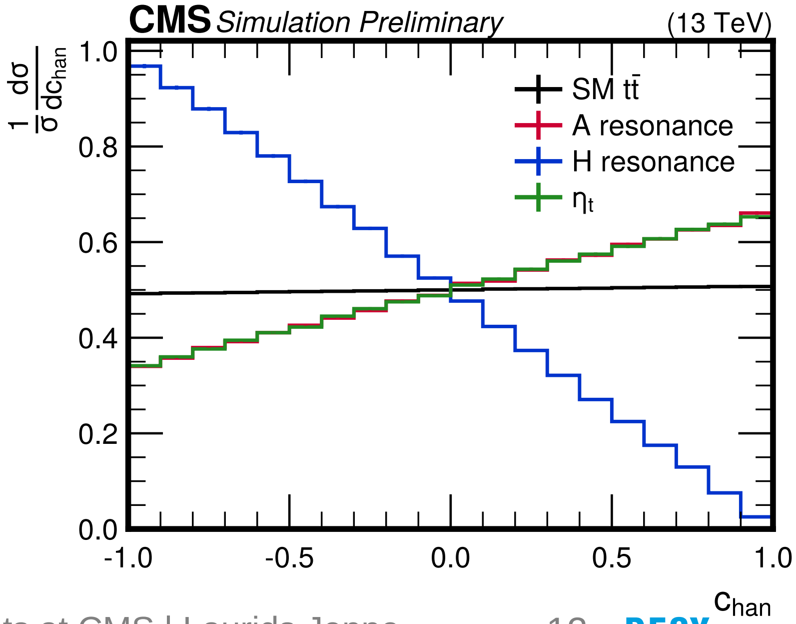
Spin correlation observables

- Both A/H and η_t predict $t\bar{t}$ production in a pure $t\bar{t}$ spin state: $^1S_0$ or $^3P_0$ (from A / η_t resp. H)
- Top decays before hadronization $\rightarrow$ transfer spin information to decay products
- Construct spin correlation observables from tops & leptons
- Variable #1: c_{hel}
  - Boost leptons into rest frames of their parent tops $\rightarrow$ Scalar product between directions of flight
  - Straight line with slope sensitive to $t\bar{t}$ spin state ("D")
  - Maximal for $^1S_0$ (from A / η_t) – separates from SM



Spin correlation observables

- Both A/H and η_t predict $t\bar{t}$ production in a pure $t\bar{t}$ spin state: $^1S_0$ or $^3P_0$ (from A / η_t resp. H)
- Top decays before hadronization $\rightarrow$ transfer spin information to decay products
- Construct spin correlation observables from tops & leptons



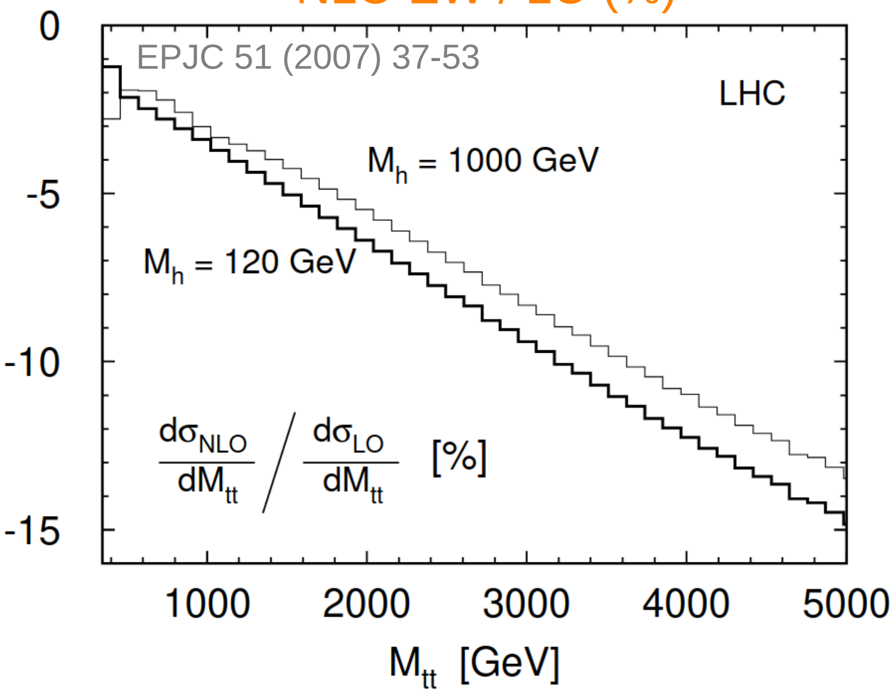
- Variable #2: c_{han}
  - Similar as c_{hel}, separating scalars from SM
  - Maximally negative slope for $^3P_0$ state (from H)
  - Construct similarly from lepton momenta, with sign flip for component parallel to top momentum

➔ 3 search variables in dilepton: m_{tt} x c_{hel} x c_{han}

Background modeling

- Major irreducible background: SM $t\bar{t}$
  - Model from NLO MC (Powheg+Pythia)
  - Correct to NNLO QCD and NLO EW from fixed-order predictions by reweighting in 2D bins of m_{tt} and cosθ*
    (EPJC 78 (2018) 537, EPJC 51 (2007) 37)
  - Normalize to NNLO+NNLL cross section
    (CPC 185 (2014) 2930)

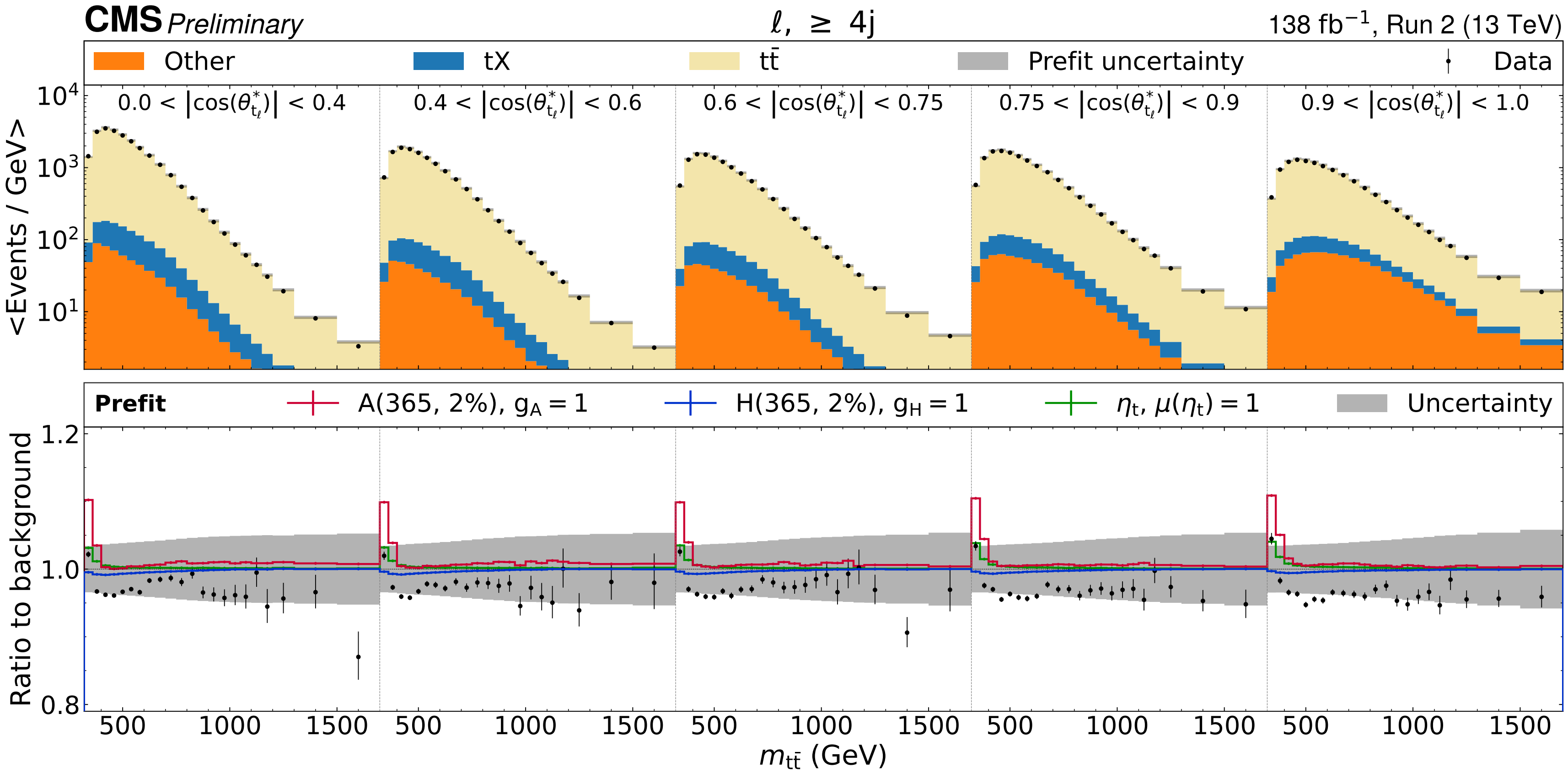
NLO EW / LO (%)



- Other backgrounds: tW, t channel single-top, rare processes (from MC)
- Z+jets in ℓℓ: from MC with data-driven normalization from Z peak sideband
- QCD+EW processes in ℓ+jets: data-driven shape from sideband with no b tags

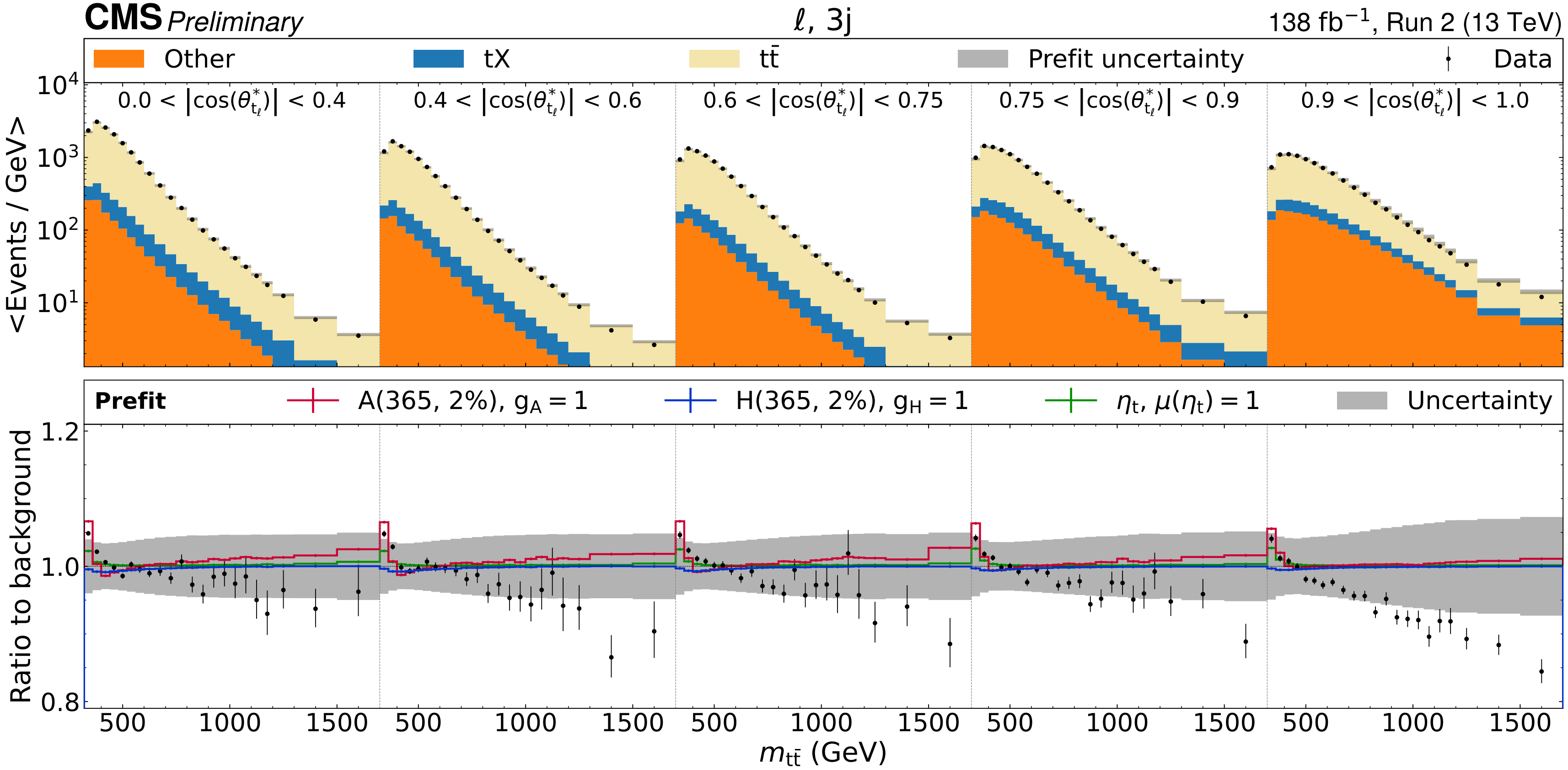
Prefit distributions: ℓ + ≥4 jets

Differences between data and prediction observed in low m_{tt} bins!



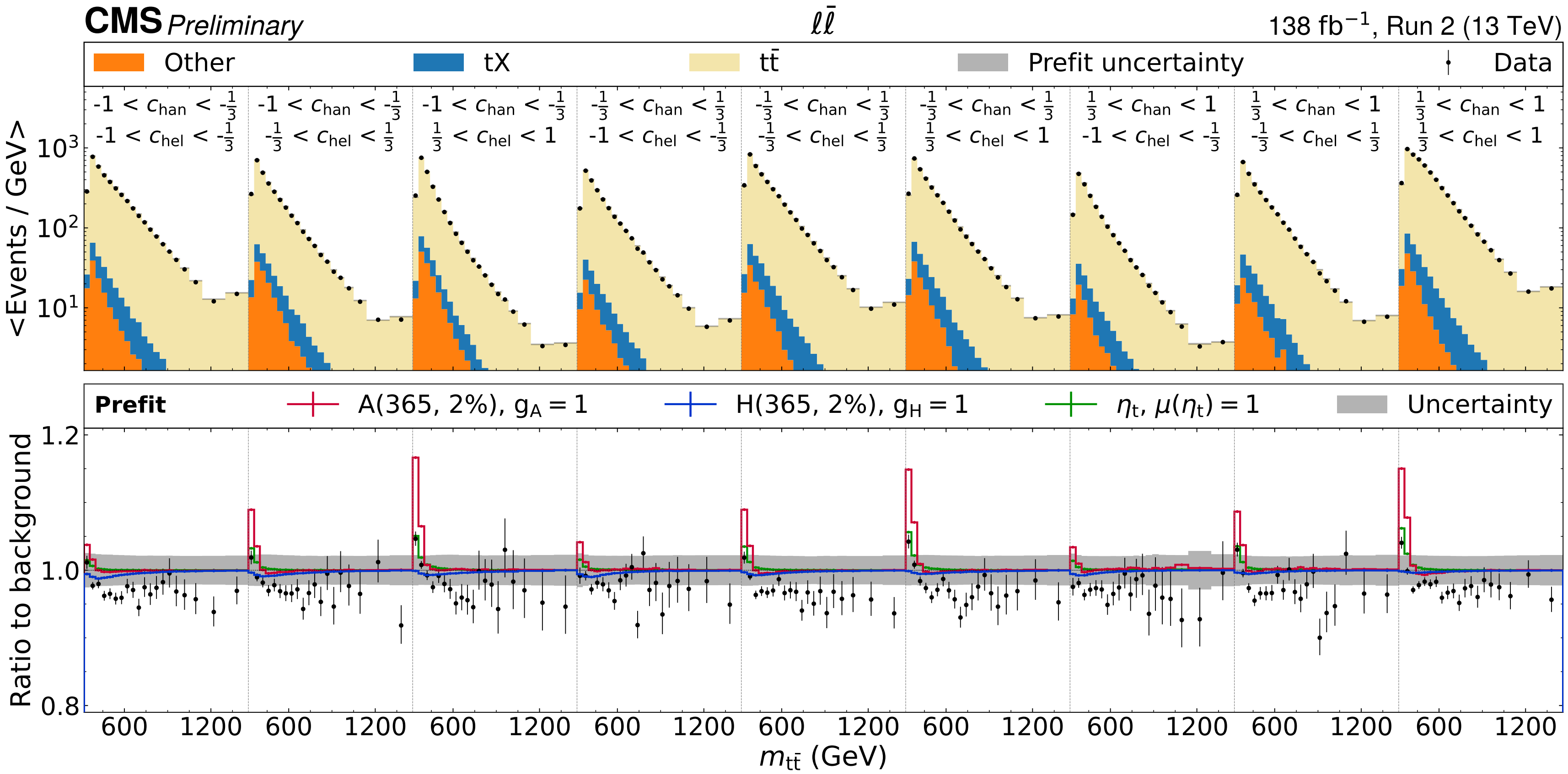
Prefit distributions: ℓ + 3 jets

Differences between data and prediction observed in low m_{tt} bins!



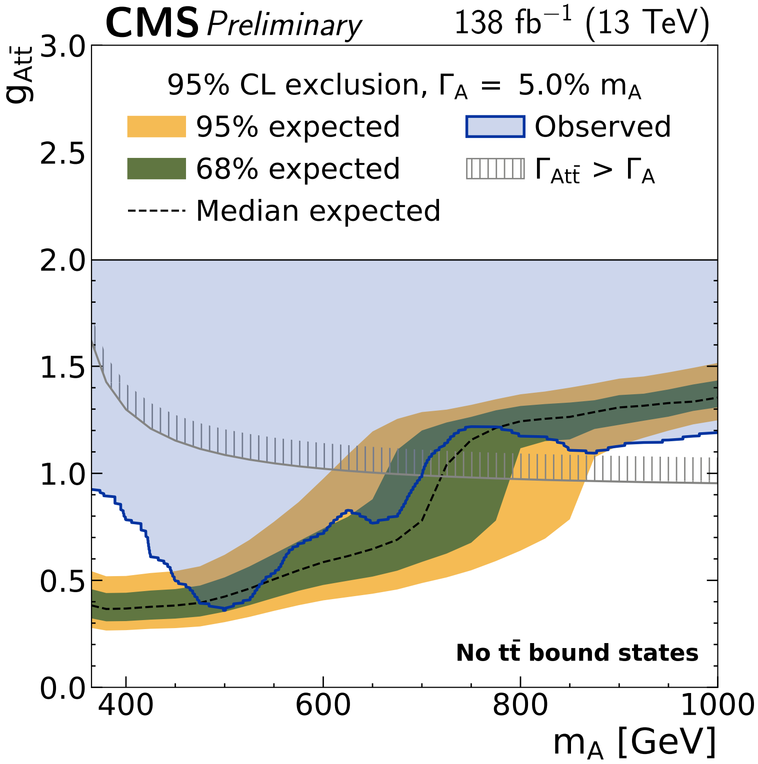
Prefit distributions: $\ell\ell$

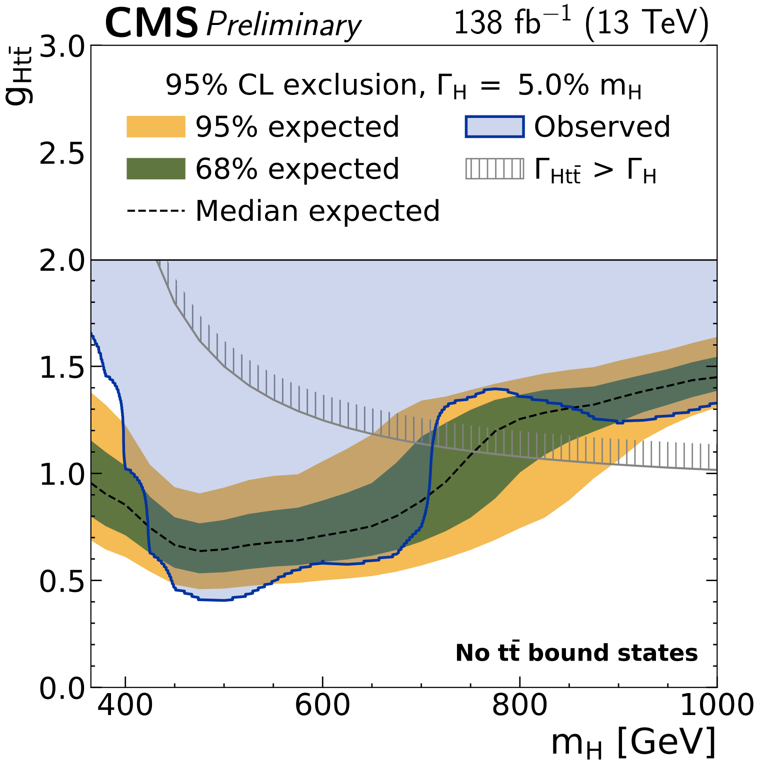
Differences between data and prediction observed in low m_{tt} bins!



A/H interpretation

- Limits on A or H using only the perturbative QCD+EW background model
- Excess at low m_{tt} visible at low A/H masses – stronger for A





A/H width: 5%

$t\bar{t}$ bound state?

- Excess is located at low $m_{t\bar{t}}$, stronger for pseudoscalar
  $\rightarrow$ could this be interpreted as $t\bar{t}$ bound state effects?
- Extract cross section using the η_t color-singlet model
  - "cross section" = difference to perturbative prediction

$$\sigma(\eta_t) = 7.1 \pm 0.8\,\text{pb}$$

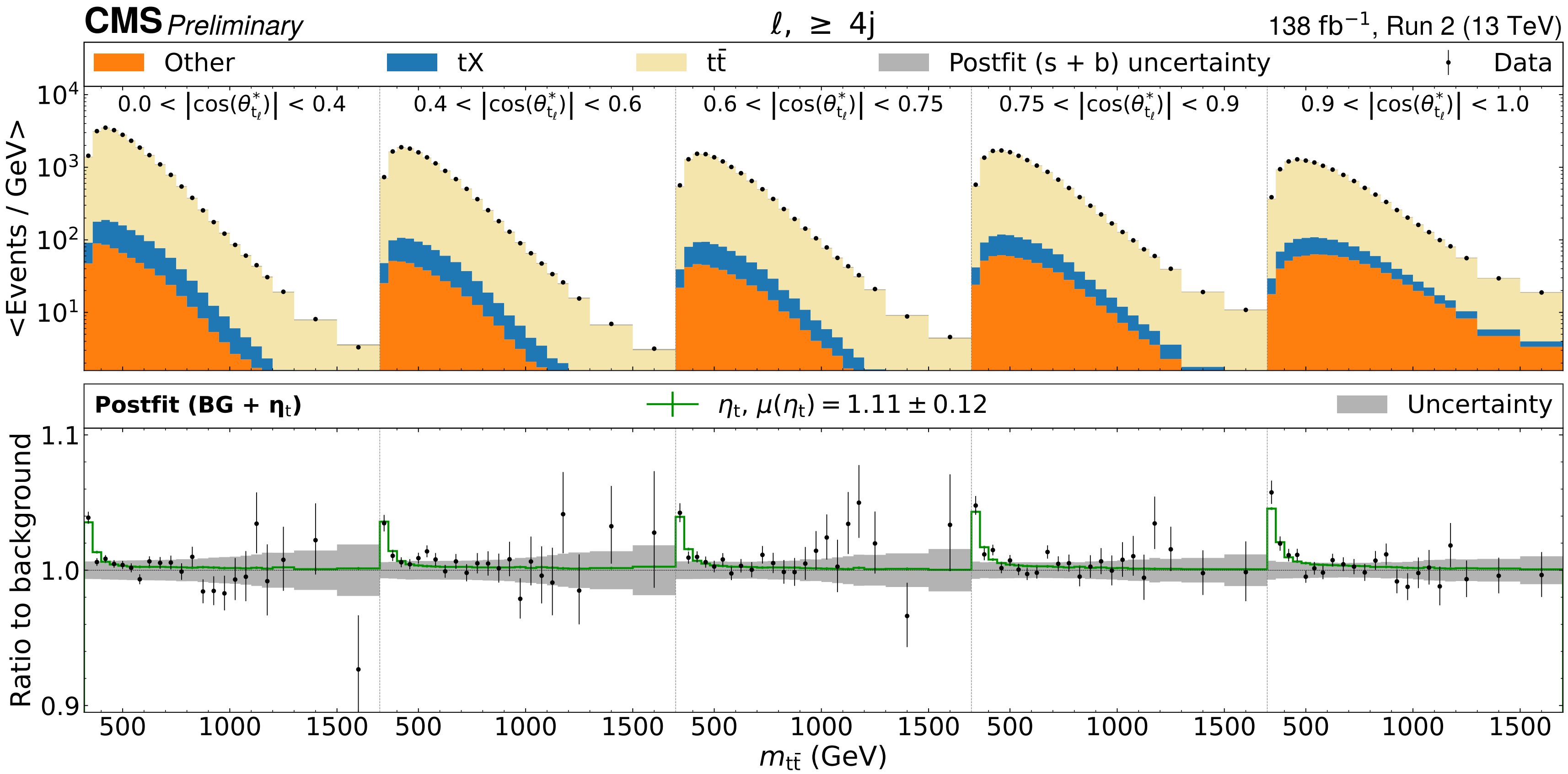
- Agrees with NRQCD prediction: $\sigma(\eta_t)^{\text{pred}} = 6.43\,\text{pb}$

(PRD 104 (2021) 034023)
(JHEP 09 (2010) 034)

- Word of caution: this model is not a complete description of a $t\bar{t}$ bound state!
  - missing e.g. color-octet states - but these are expected to be small

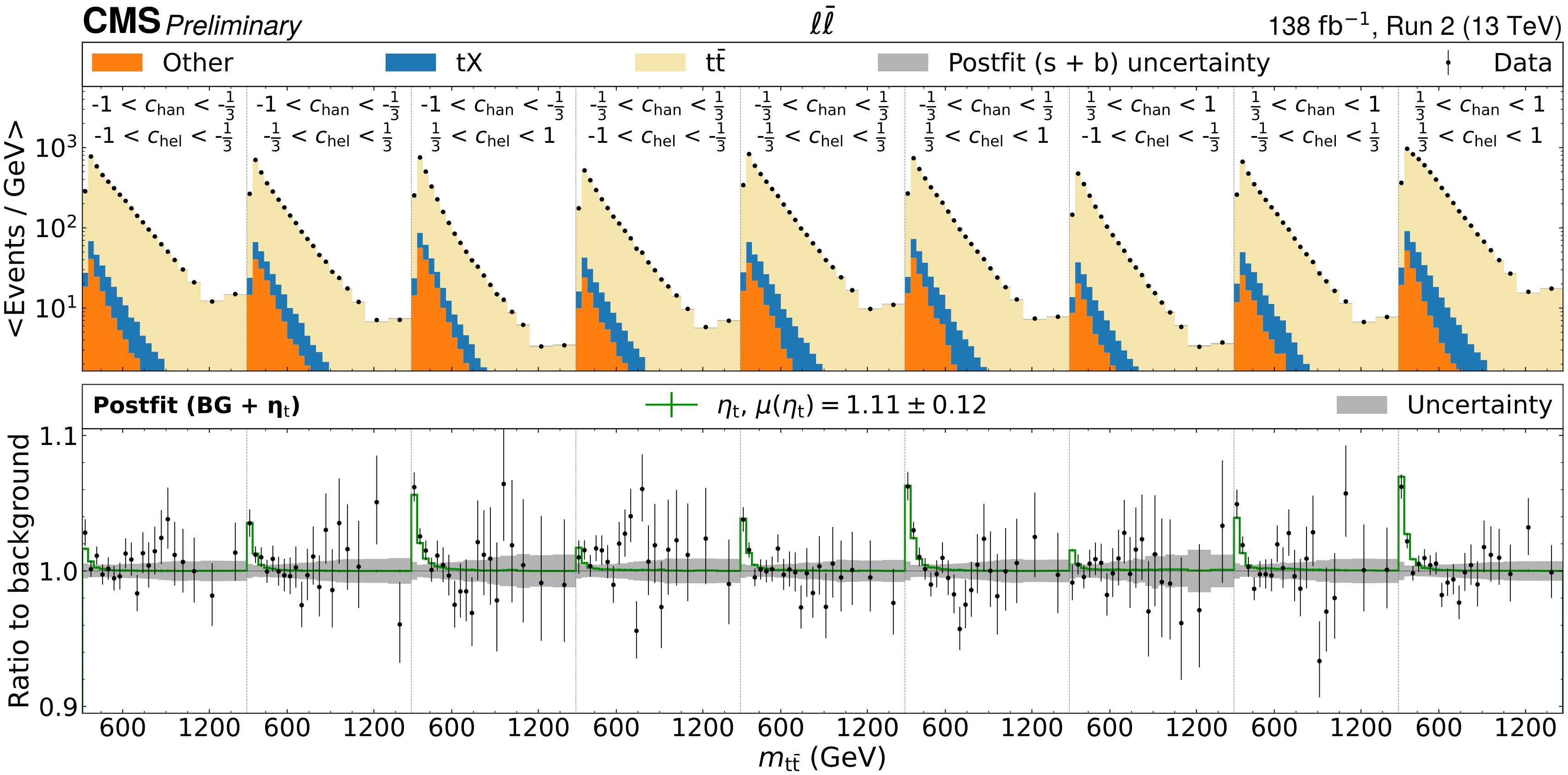
Postfit distributions: η_t (ℓ + ≥4 jets)

Postfit for η_t model describing the data well



Postfit distributions: η_t ($\ell\ell$)

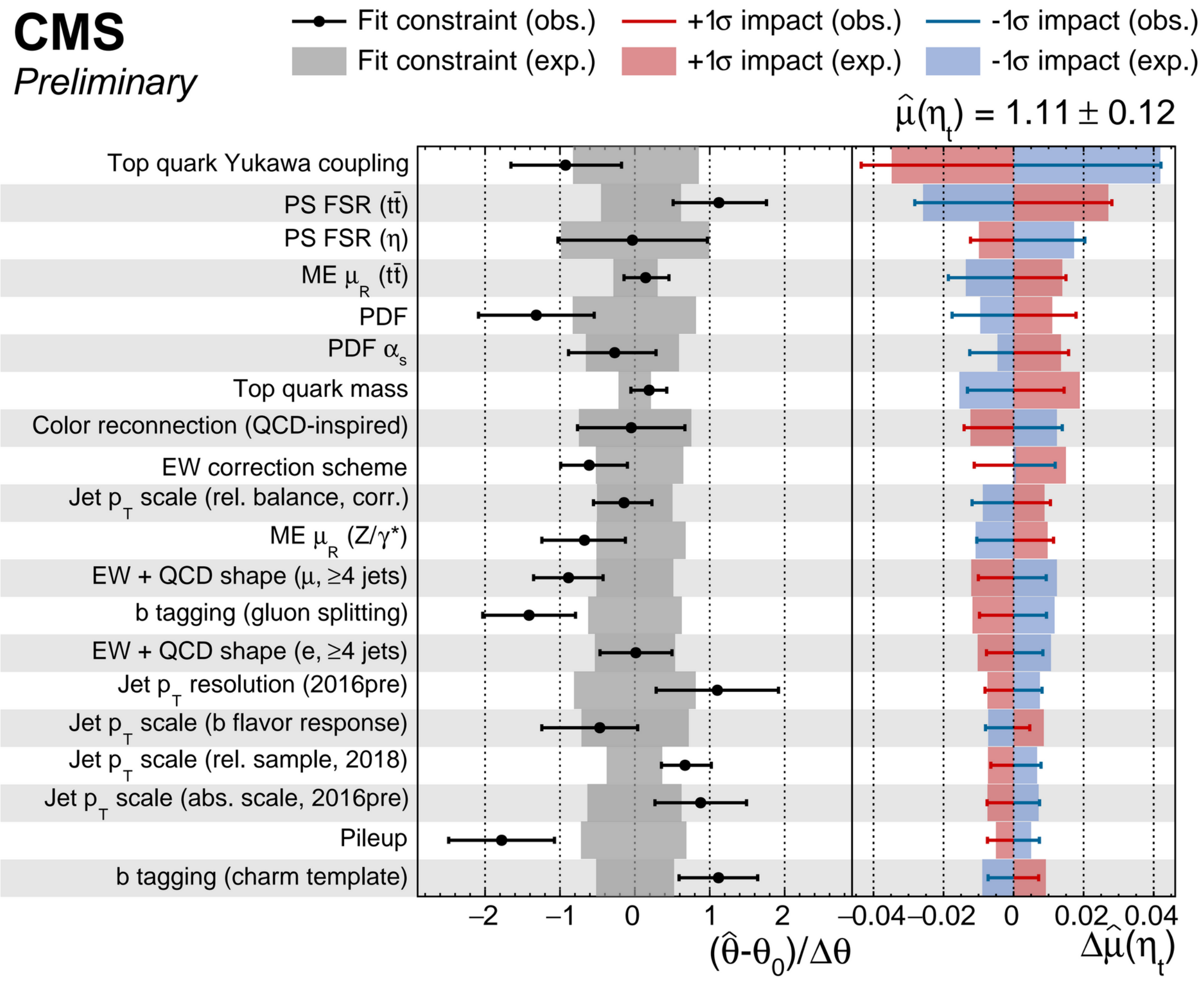
Postfit for η_t model describing the data well



Uncertainties

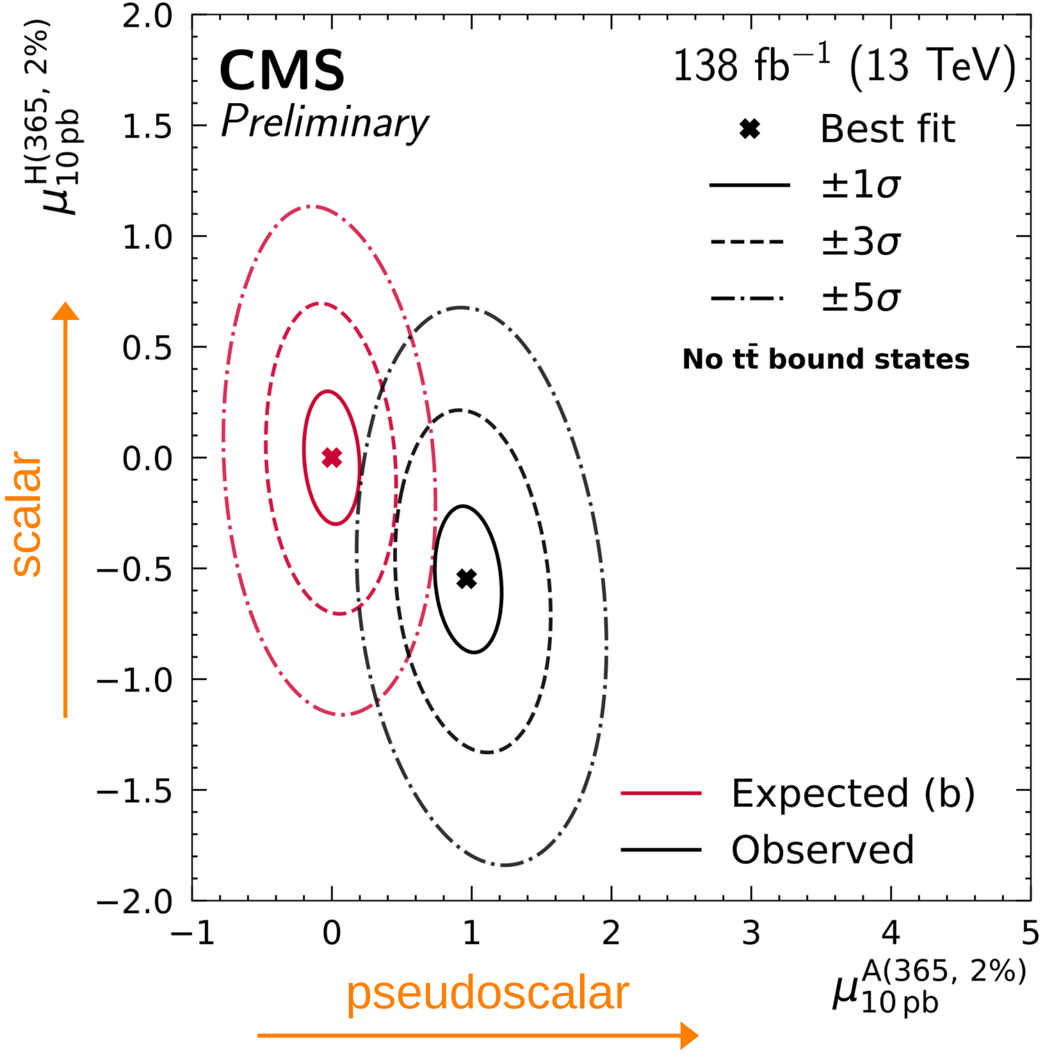
- Uncertainty on η_t cross section dominated by background modeling

- Leading systematic sources:
  - EW corrections, including SM Top-Higgs Yukawa:
    $y_t = 1.00^{+0.11}_{-0.12}$ (EPJC 79 (2019) 421)
  - Parton shower scales
  - Missing higher orders
  - PDF
  - Top mass



Parity of the excess

- Can we quantify whether the excess is scalar or pseudoscalar?
- Take low-mass A/H resonances as proxies for pure $^1S_0$ and $^3P_0$ $t\bar{t}$ states
- ➔ 2D fit with arbitrary signal strengths
- Data prefers pure $^1S_0$ / pseudoscalar
- scalar component compatible with 0 at the level of ~ 2 SD

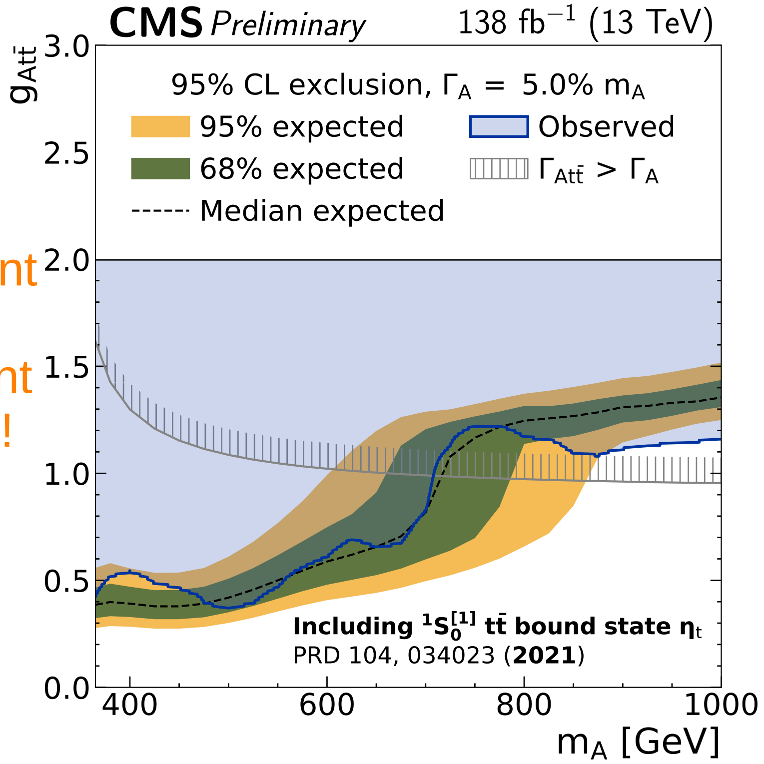


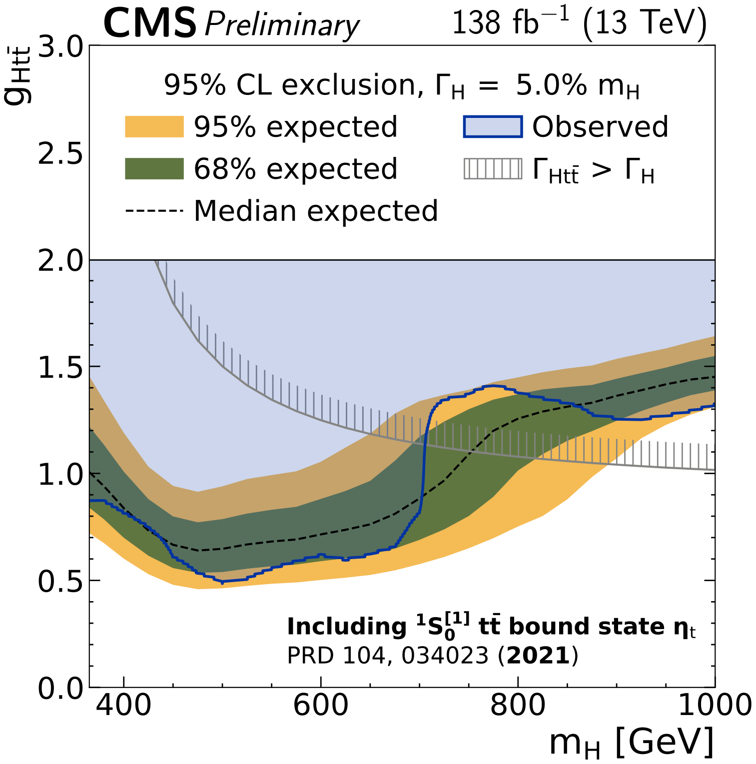
A/H limits including η_t

- QCD + η_t describes data well → set (BSM) A/H limits
  → η_t added as an additional BG process with free-floating normalization

Excess no longer present

Most stringent limits to date!





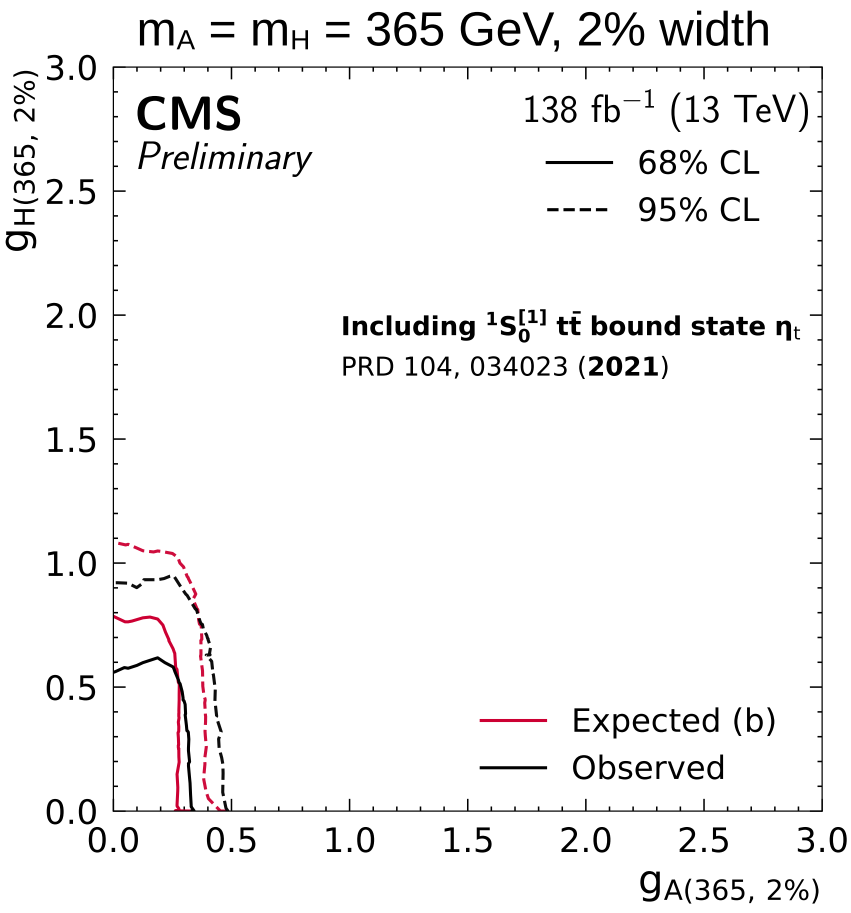
A/H width: 5%

A+H interpretation

- BSM models (e.g. 2HDM) often predict the simultaneous presence of A and H

➔ Model-independent exclusion contours for both A and H couplings
  - numerical Feldman-Cousins method

- Input for bounds on concrete BSM models



Summary

- Search for new spin-0 (pseudo)scalars in $t\bar{t}$ final states with full Run 2 dataset
- Dilepton and lepton+jets channels, using m_{tt}, angular and spin observables
- Observed excess in data at low m_{tt} – consistent with pseudoscalar
  - Interpretations in terms of a simplified model of a $t\bar{t}$ bound state η_t or a generic pseudoscalar A and scalar H
  - Extracted cross section for a specific η_t model (PRD 104 (2021) 034023)
- Set stringent limits on A, H, and A+H with η_t included in the background

- For the future:
  A complete non-relativistic QCD calculation of $t\bar{t}$ bound state effects is crucial!
  Input from theory needed

Reference: **CMS-PAS-HIG-22-013**

Backup

Comparison with bb4l

- Our nominal $t\bar{t}$ MC (Powheg hvq+Pythia) includes off-shell top quarks only approximately (narrow-width approximation)
  but η_t lies below the threshold – off-shell!
- Full off-shell at NLO: `bb4l`
  generates $pp \to b\bar{b}\ell\bar{\ell}\nu\bar{\nu}$ using the full amplitude
  for now, only available in dilepton

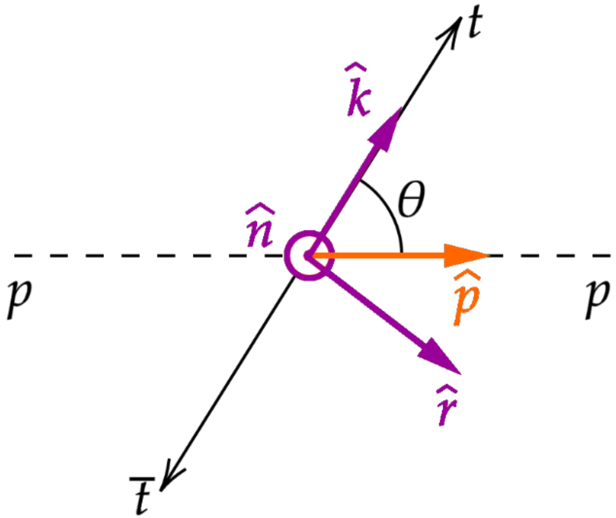
- Redo our extraction with `bb4l` for the $t\bar{t}$+tW prediction in $\ell\ell$ only

➔ Results compatible – excess clearly present also with `bb4l`

Definition of c_{hel} and c_{han}

- Start in $t\bar{t}$ rest frame, boost leptons into rest frames of their parent tops
- Define lepton three-momenta $\hat{\ell}^+$ and $\hat{\ell}^-$ w.r.t $\{\hat{k}, \hat{r}, \hat{n}\}$ basis:
  - $\hat{k}$: direction of flight of the top quark
  - $\hat{r}$: orthogonal to $\hat{k}$ in the scattering plane
  - $\hat{n}$: orthogonal to $\hat{k}$ and $\hat{r}$

$$c_{\mathrm{hel}} = -(\hat{\ell}^+)_k(\hat{\ell}^-)_k - (\hat{\ell}^+)_r(\hat{\ell}^-)_r - (\hat{\ell}^+)_n(\hat{\ell}^-)_n$$
$$c_{\mathrm{han}} = +(\hat{\ell}^+)_k(\hat{\ell}^-)_k - (\hat{\ell}^+)_r(\hat{\ell}^-)_r - (\hat{\ell}^+)_n(\hat{\ell}^-)_n$$



- It can be shown that they follow a straight line with

$$\frac{1}{\sigma}\frac{d\sigma}{dc_{\mathrm{hel}}} = \frac{1}{2}(1 - D\, c_{\mathrm{hel}}) \qquad \frac{1}{\sigma}\frac{d\sigma}{dc_{\mathrm{han}}} = \frac{1}{2}\left(1 + D^{(k)}\, c_{\mathrm{han}}\right)$$

(JHEP 03 (2024) 099)

List of systematic uncertainties

Experimental

- Jet energy corrections - split into 11 subsources
- Jet energy resolution
- Unclustered p^T_{miss} (uncorrelated between years)
- Luminosity – correlated and decorrelated parts between years
- Pileup
- Trigger efficiencies (separate for ℓℓ / ℓj)
- Electron efficiencies (reco. & ID)
- Muon efficiencies – split into syst. and stat.
- B tagging and mistagging efficiencies
  - B tagging split into subsources
- L1 ECAL prefiring (where applicable)
- Data-driven EW+QCD BG (ℓ+jets) : shape & rate (50%) uncorrelated between channels
- Data-driven Z+jets normalization (ℓℓ)

Theory

- Factorization & renormalization scales:
  - $t\bar{t}$, tW, tq, Z+jets; η_t (BG or signal), A/H signal
  - Uncorrelated between processes
  - $t\bar{t}$: including cross section variation
- Same for initial & final state radiation PS scales
- MC top mass: ±1GeV (interpolated from ±3GeV)
  - Also including cross section variations
- ME-PS matching (h_{damp})
- Underlying event tune
- Color reconnection: 3 different samples
- PDF: PCA performed on final templates from 100 replicas → only leading component considered
- PDF α_s
- Electroweak corrections:
  - SM Higgs-Top Yukawa coupling (1 +0.11 -0.12)
  - EW correction scheme (additive v. multiplicative)
- Minor BG cross sections: 15% for tW and tq; 30% for Diboson and $t\bar{t}$+X

List of MC generators

Process	QCD order	ME Generator
$t\bar{t}$	NLO	POWHEG v2 (hvq)
tW	NLO	POWHEG v2 (ST_wtch)
Z+jets	NNLO	POWHEG v2 (Zj MiNNLO)
t-channel single top	NLO	POWHEG v2 (ST_tch) + MADSPIN
s-channel single top	NLO	MG5_aMC@NLO
$t\bar{t}W$	NLO	MG5_aMC@NLO
$t\bar{t}Z$	NLO	MG5_aMC@NLO
WW, WZ & ZZ	LO	PYTHIA 8.2
A/H signal	LO	MG5_aMC@NLO
η_t signal	LO	MG5_aMC@NLO

Data-driven Z+jets normalization

- b jets in Z+jets are known to be badly modeled in MC – might lead to wrong normalization after requiring >= 1 btag
- Take normalization from Z peak sideband ($R_{in/out}$ method)
- Use weaker assumption than standard $R_{in/out}$ (“ratio of ratios”):
  Get $R_{in/out}$ in 0 b tag sideband; take “ratio of ratios” for ≥ 1 and 0 btags from MC

$$\frac{(R^{\geq 1b}_{in/out})_{data}}{(R^{\geq 1b}_{in/out})_{MC}} = \frac{(R^{0b}_{in/out})_{data}}{(R^{0b}_{in/out})_{MC}} \quad \rightarrow \quad \mathrm{SF} = \frac{(N^{\geq 1b}_{out})_{data}}{(N^{\geq 1b}_{out})_{MC}} = \frac{(N^{\geq 1b}_{in})_{data}}{(N^{\geq 1b}_{in})_{MC}} \frac{(R^{0b}_{in/out})_{MC}}{(R^{0b}_{in/out})_{data}}.$$

with $N_{data} = N^{\ell\ell}_{data} - 0.5 N^{e\mu}_{data} k_{\ell\ell}, \quad \text{where} \;\; k_{ee} = \frac{1}{k_{\mu\mu}} = \sqrt{\frac{N^{ee}_{data}}{N^{\mu\mu}_{data}}}$

EW corrections to $t\bar{t}$

- Our EW correction (Hathor) is NLO in EW but LO in QCD
- Ambiguity on how to apply EW corrections to (N)NLO simulation
- Nominal choice: multiplicative

Hathor

Powheg

$$\sigma^{\text{rew.}} = \sigma^{\text{LO EW}}_{\text{NLO QCD}} \times \frac{\sigma^{\text{NLO EW}}_{\text{LO QCD}}}{\sigma^{\text{LO EW}}_{\text{LO QCD}}}$$

MadGraph

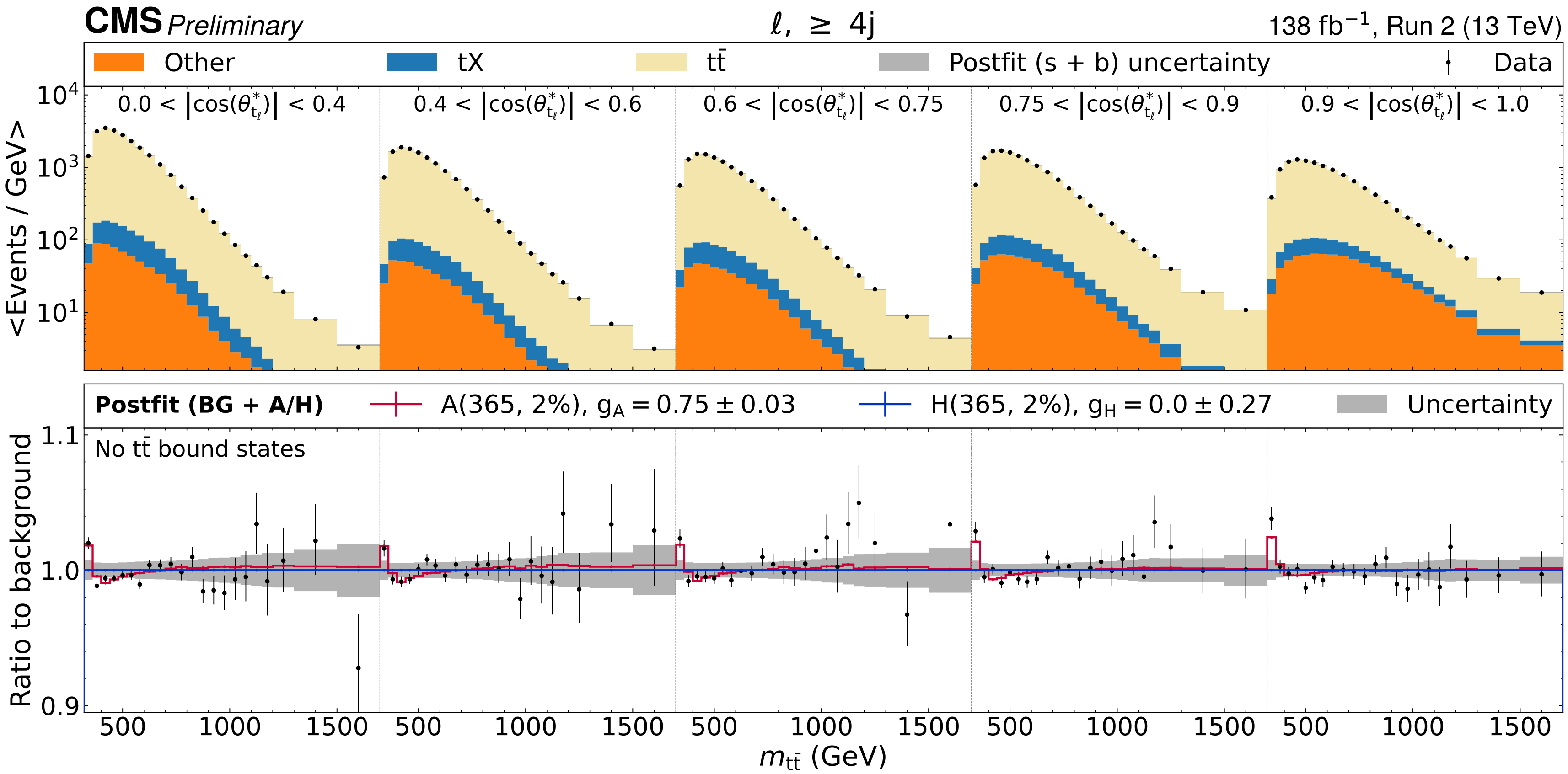
- Alternate choice: additive

$$\sigma^{\text{rew.}} = \sigma^{\text{LO EW}}_{\text{NLO QCD}} + \sigma^{\text{NLO EW}}_{\text{LO QCD}} - \sigma^{\text{LO EW}}_{\text{LO QCD}}$$

- Difference treated as systematic uncertainty

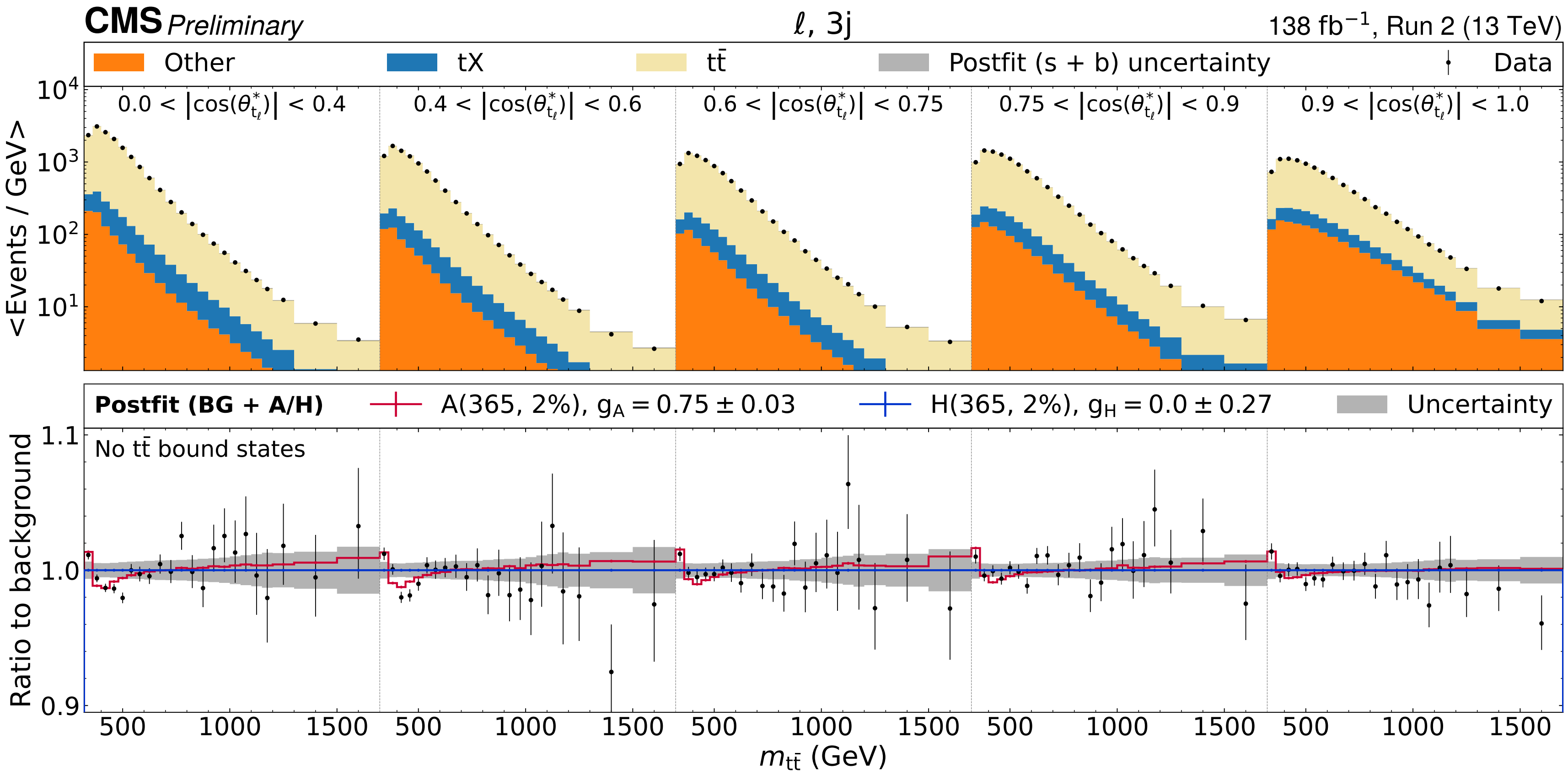
Postfit distributions: A/H

Postfit for A, 365 GeV, 2% width (best fit point)



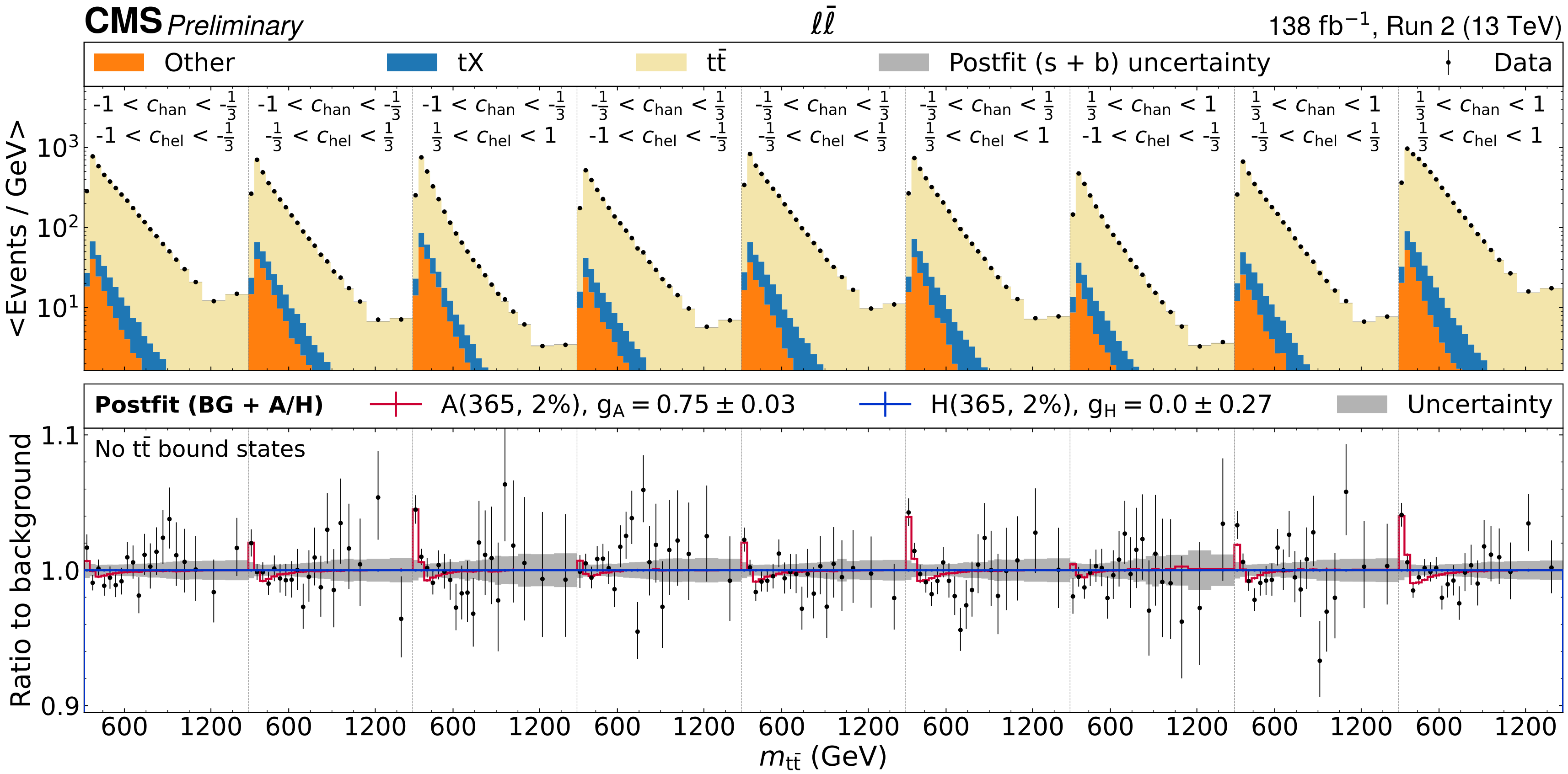
Postfit distributions: A/H

Postfit for A, 365 GeV, 2% width (best fit point)



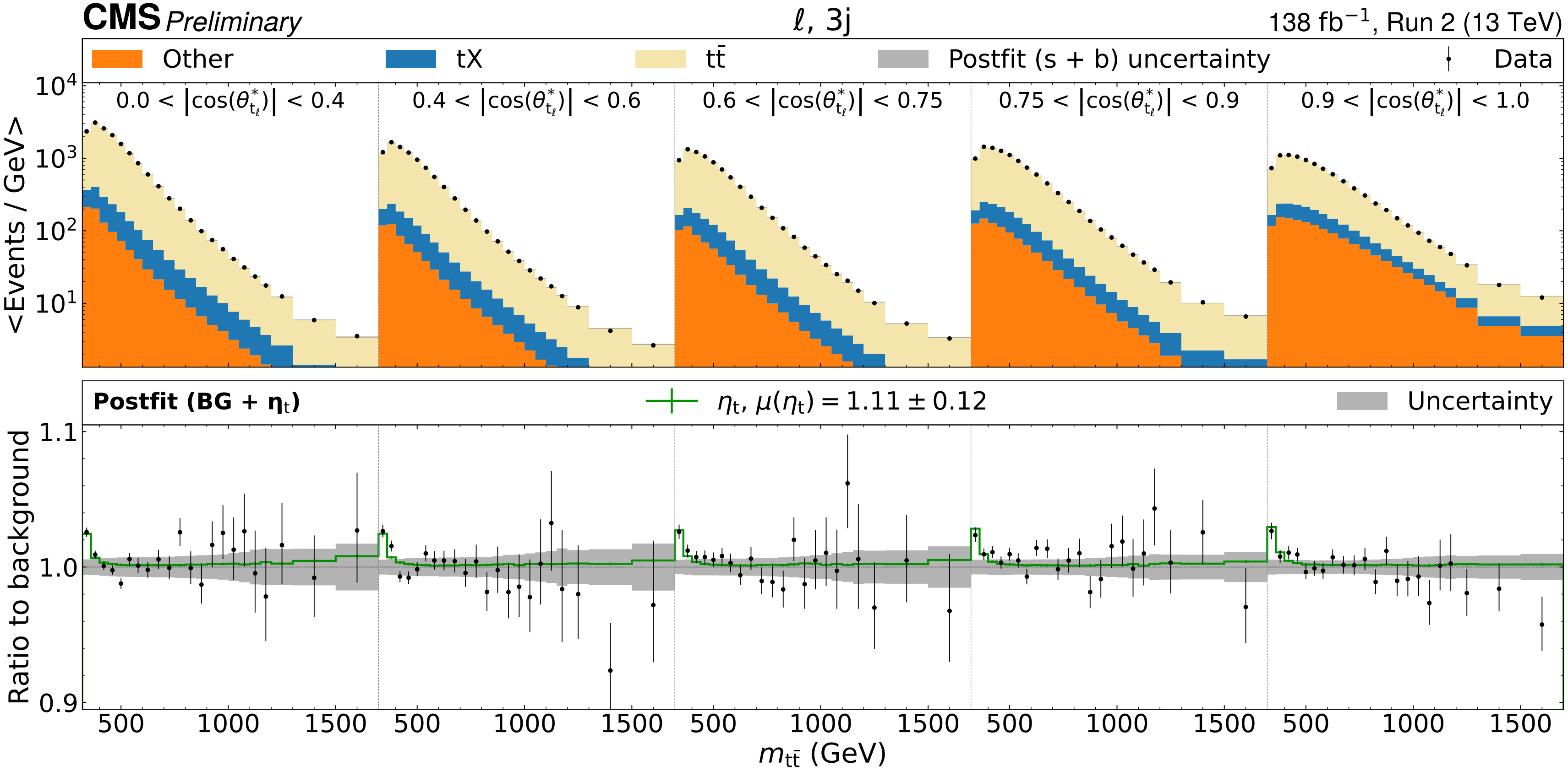
Postfit distributions: A/H

Postfit for A, 365 GeV, 2% width (best fit point)



Postfit distributions: η_t

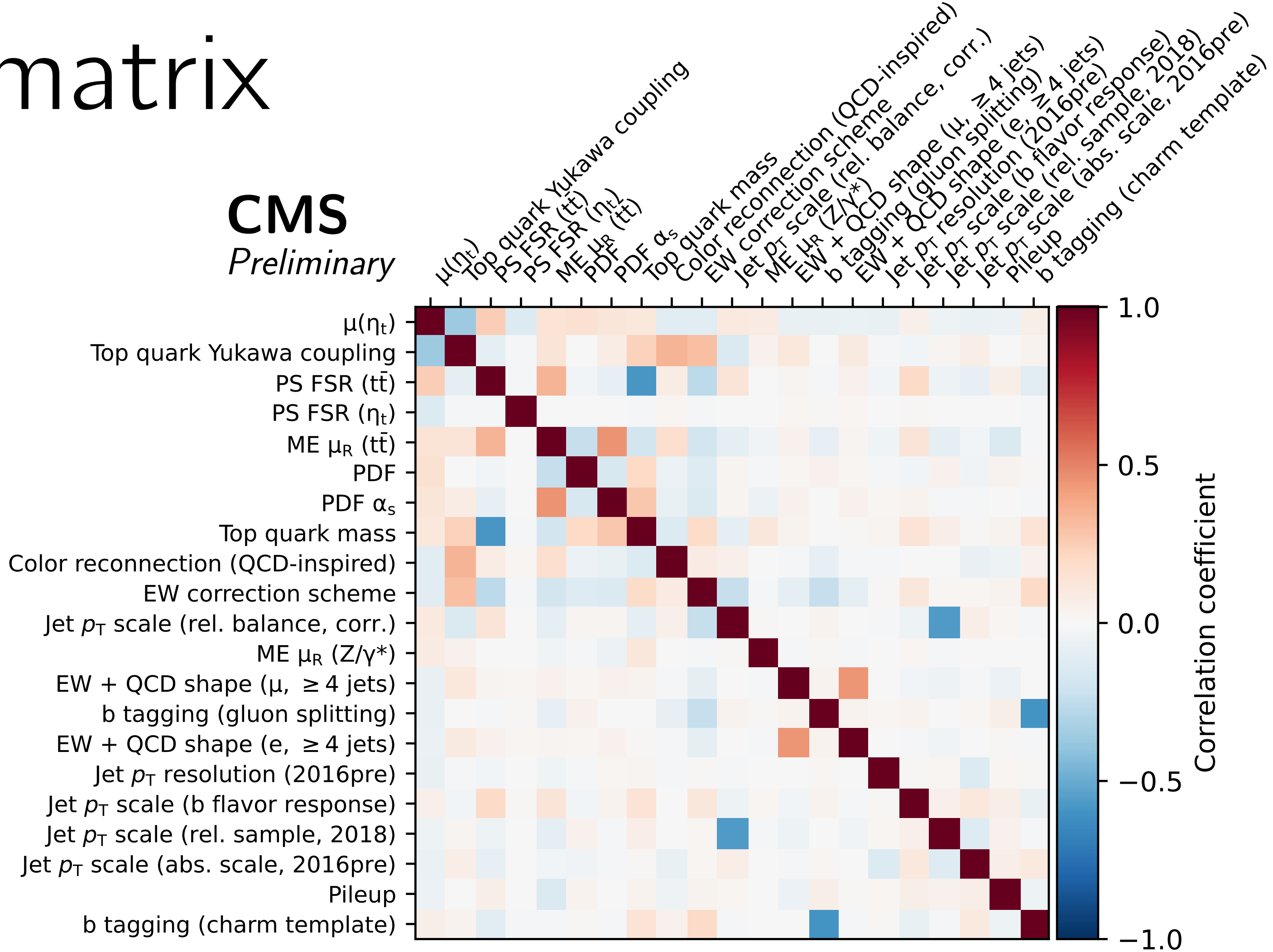
Postfit for η_t



DESY.

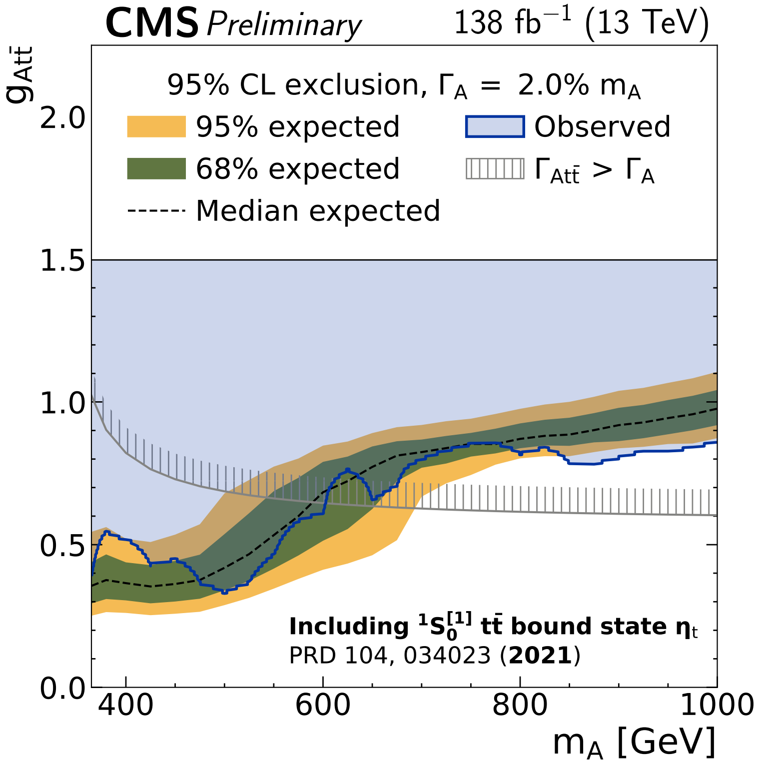
Correlation matrix

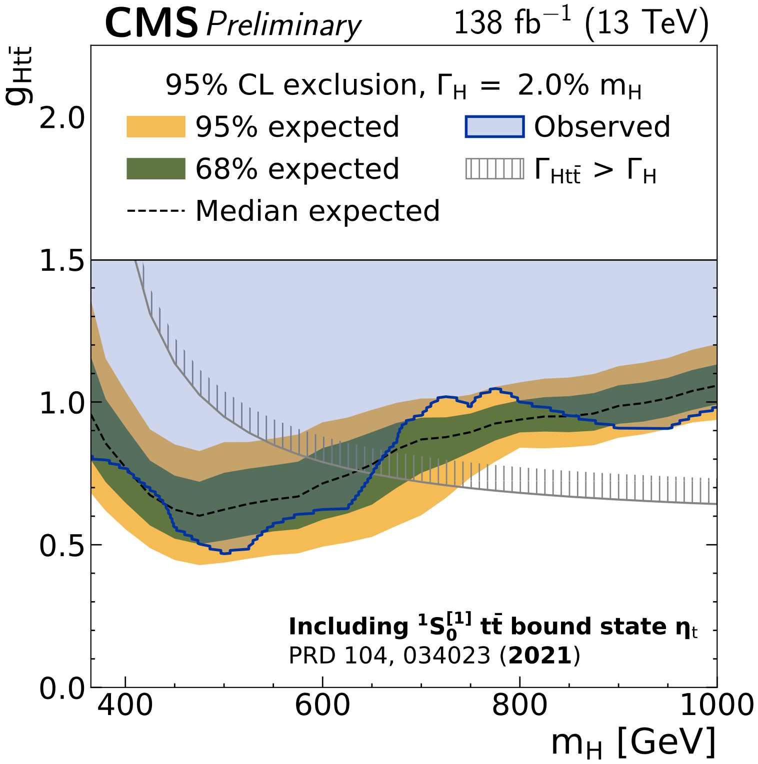
- Further assess uncertainty modeling through correlations of nuisance parameters



A/H limits including η_t

- Limits at different A/H widths

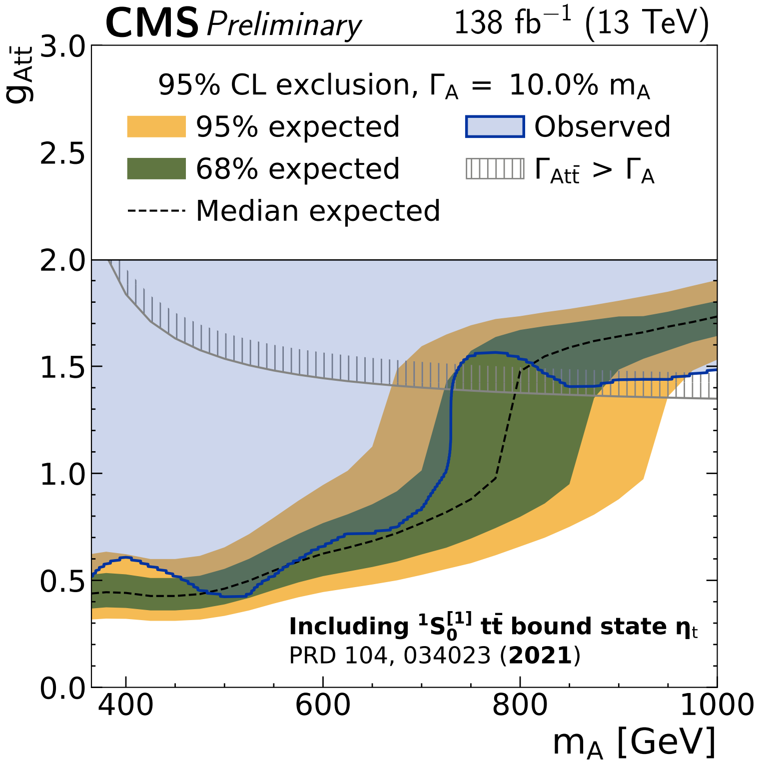


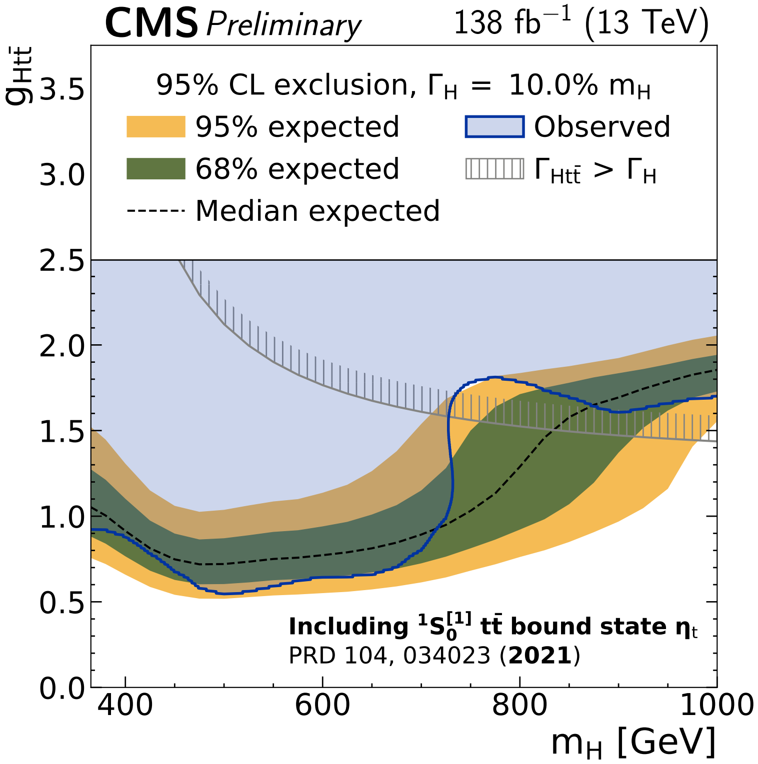


2% width

A/H limits including η_t

- Limits at different A/H widths

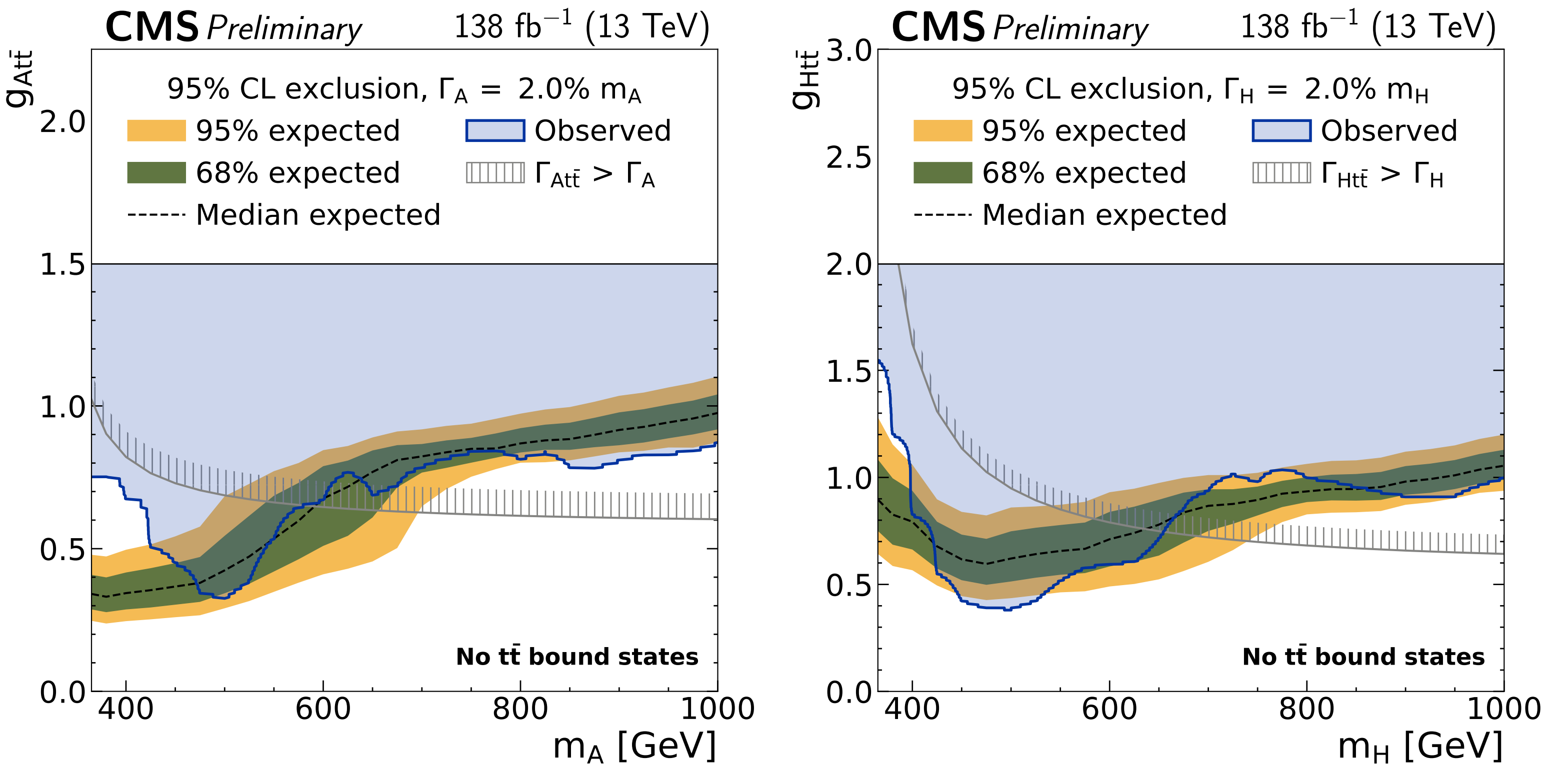




10% width

A/H limits without η_t

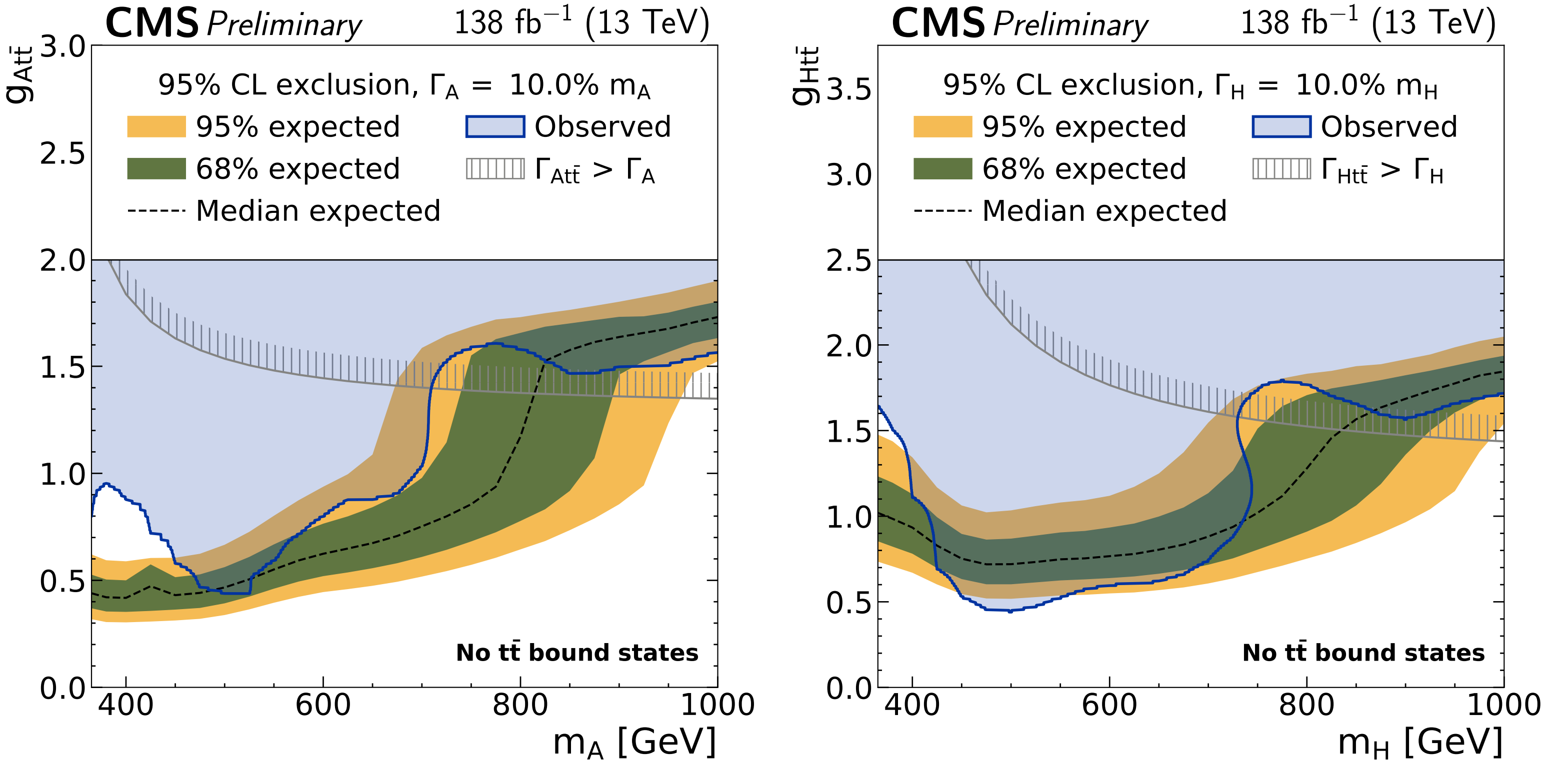
- Limits at different A/H widths for perturbative QCD background only



2% width

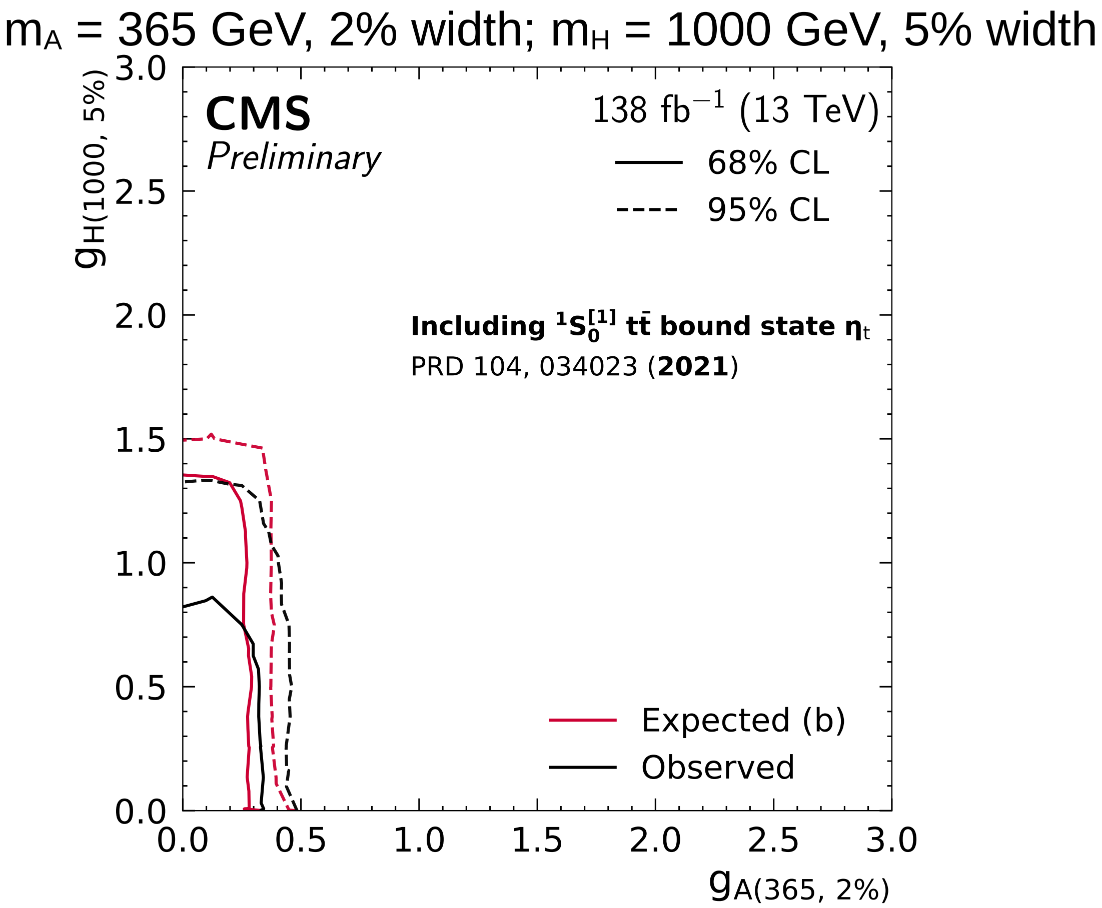
A/H limits without η_t

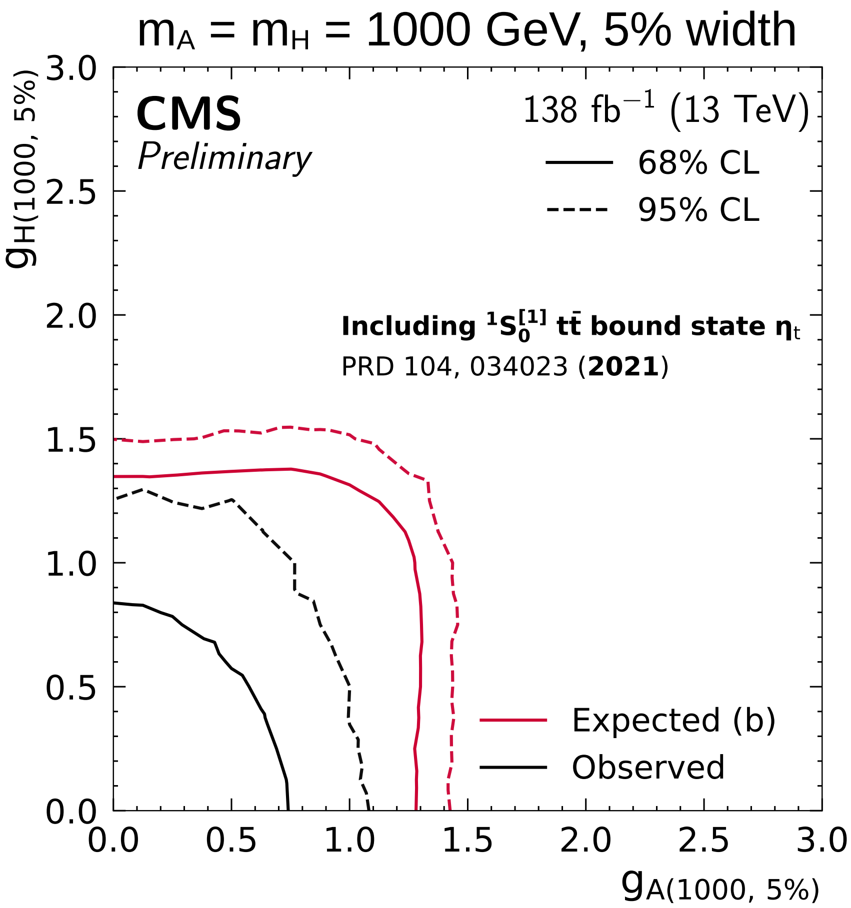
- Limits at different A/H widths for perturbative QCD background only



10% width

A+H interpretation

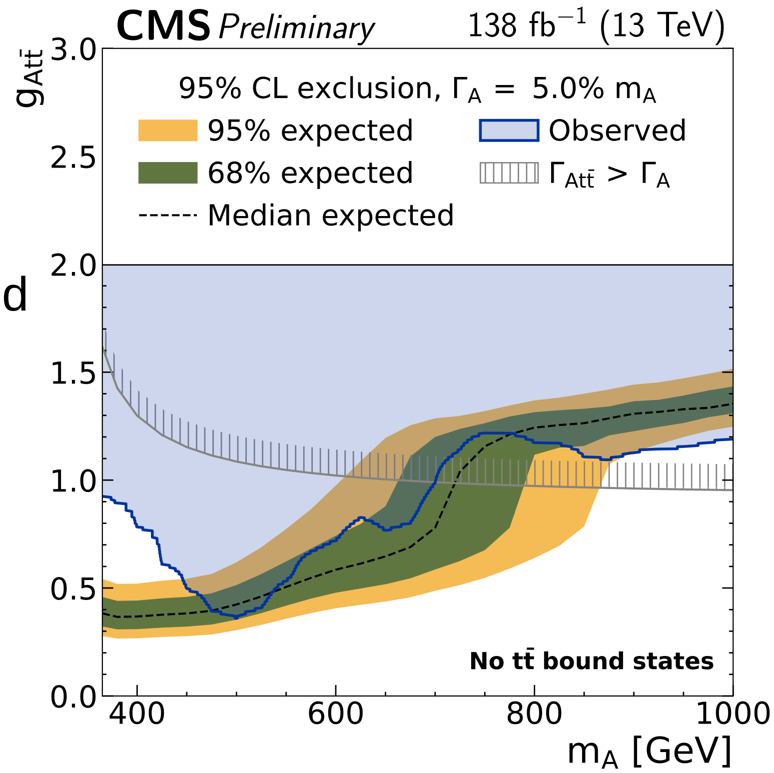




Comparison with ATLAS limits

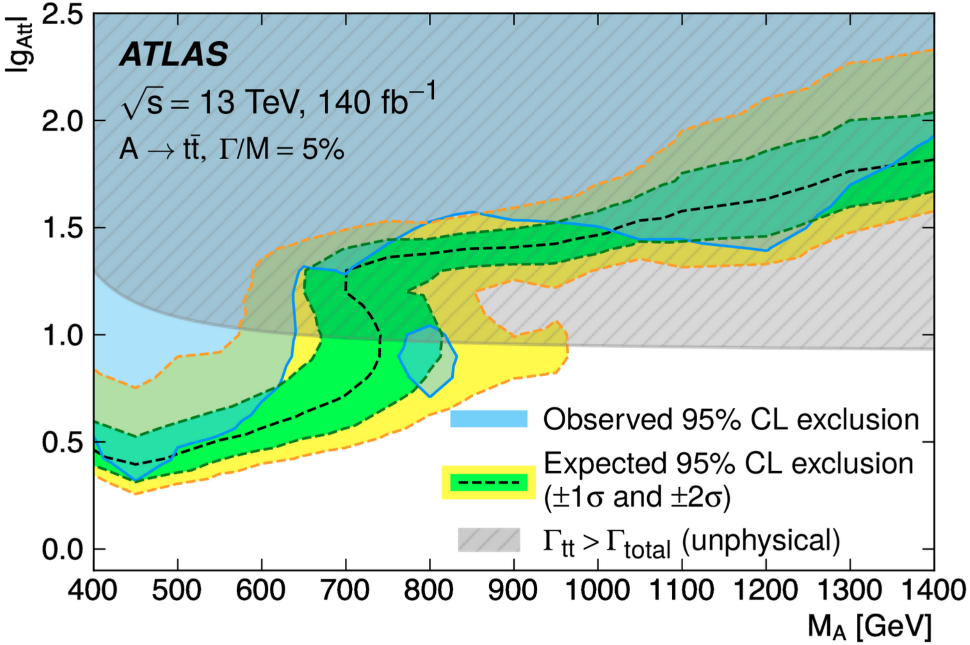
- Compare expected and observed limits with ATLAS

No η_t included for CMS



Pseudoscalar
5% width

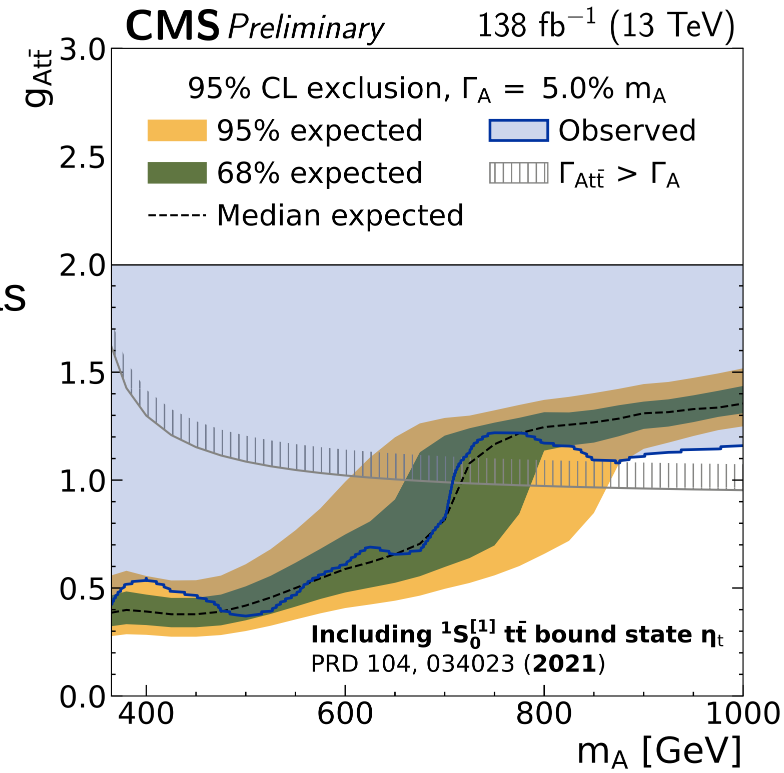
(JHEP08 (2024) 013)



Comparison with ATLAS limits

- Compare expected and observed limits with ATLAS

η_t included as background



Pseudoscalar
5% width

(JHEP08 (2024) 013)