Measurement of the inclusive $t\bar{t}$ cross section and search for additional scalars in $t\bar{t}$ final states at the CMS experiment

PhD Disputation at Hamburg University
Laurids Jeppe

13.10.2025

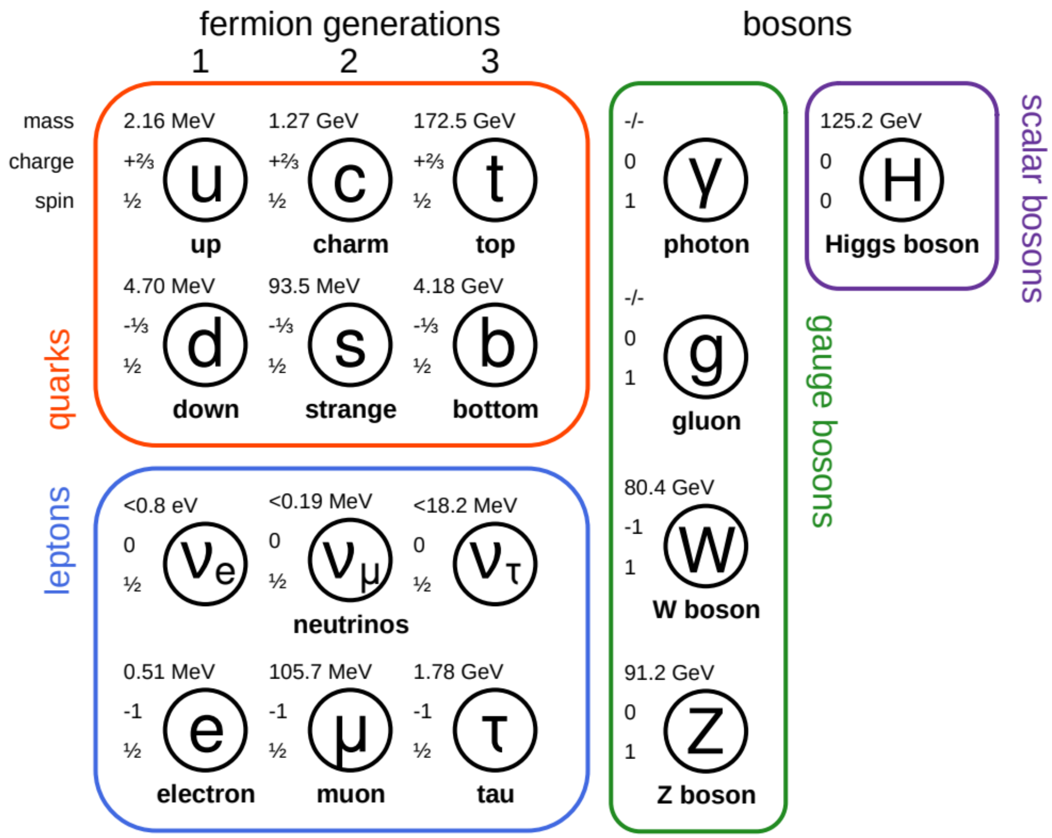
What do we do in High Energy Physics?

What do we do in High Energy Physics?

$\rightarrow$ we investigate Nature on the smallest known scales

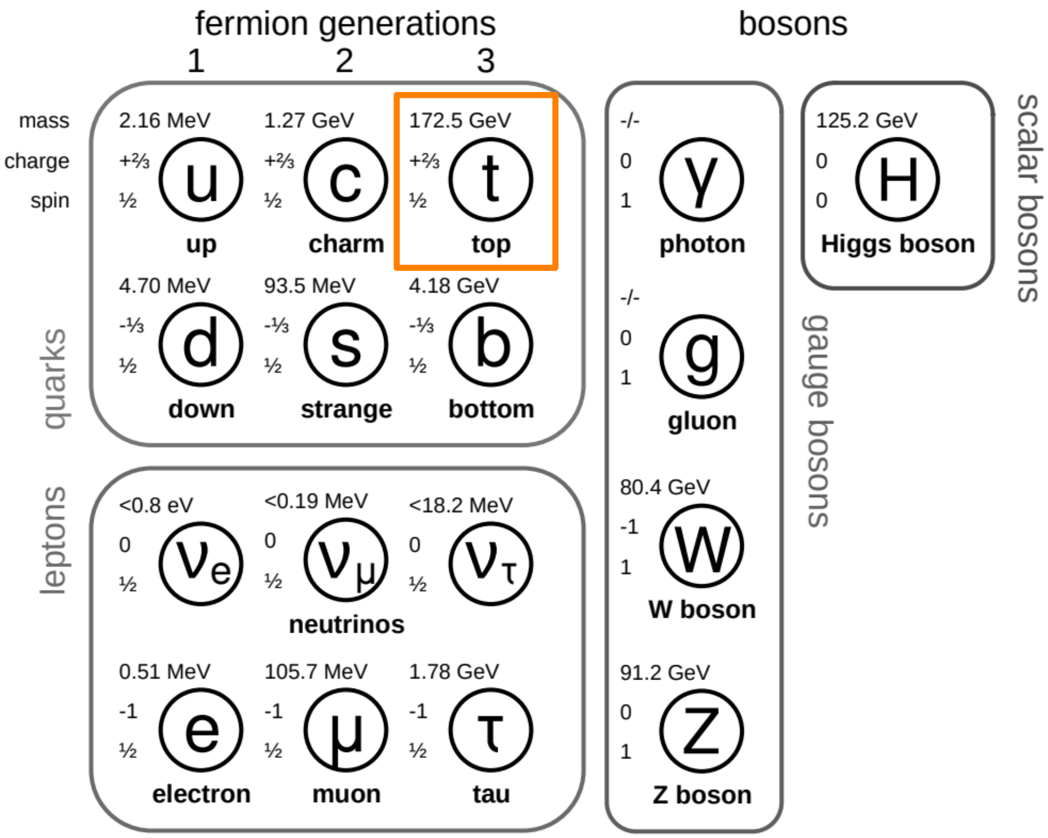
The Standard Model

- Relativistic Quantum Field Theory
- Particles = fundamental excitations of fields
- Most successful theory of our universe at small scales
- But: known to be incomplete!
  - neutrino masses, dark matter, ...



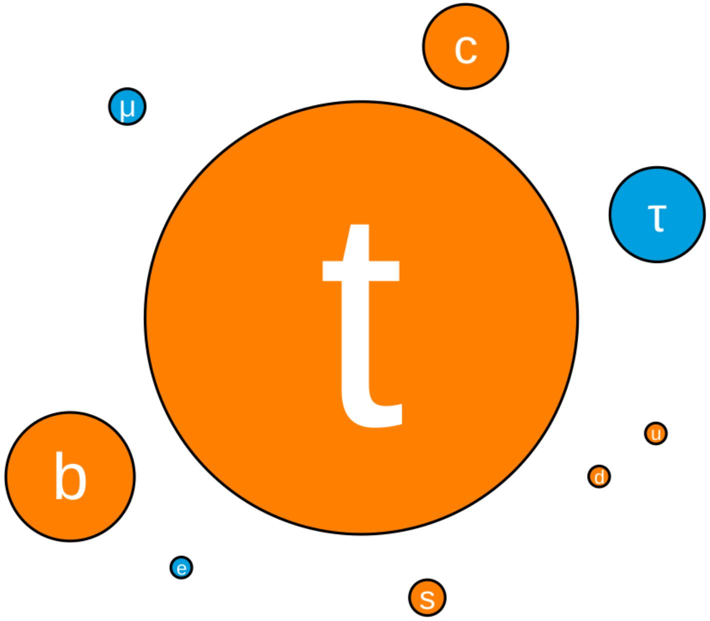
The Standard Model

- Relativistic Quantum Field Theory
- Particles = fundamental excitations of fields
- Most successful theory of our universe at small scales
- But: known to be incomplete!
  - neutrino masses, dark matter, ...



The Top Quark

- The top quark is special:
  - Heaviest elementary particle
  - Largest coupling to Higgs
  - ➔ gateway to new physics:
    e.g. additional Higgs sectors
  - Short lifetime ~ 10^{-25} s
  - No hadronization - bare quark
  - ➔ access to spin properties



- Understanding the top is crucial for SM and BSM physics!

What do we do in High Energy Physics?

$\rightarrow$ we investigate Nature on the smallest known scales

$\rightarrow$ test our theory: the Standard Model

$\rightarrow$ search for physics beyond the SM

How do we do this?

$\rightarrow$ collider experiments

The Large Hadron Collider

- The largest particle accelerator built (so far)
  - 27 km circumference
  - 8 T magnetic fields from superconducting magnets
  - 3 x 10^{14} protons in the machine simultaneously

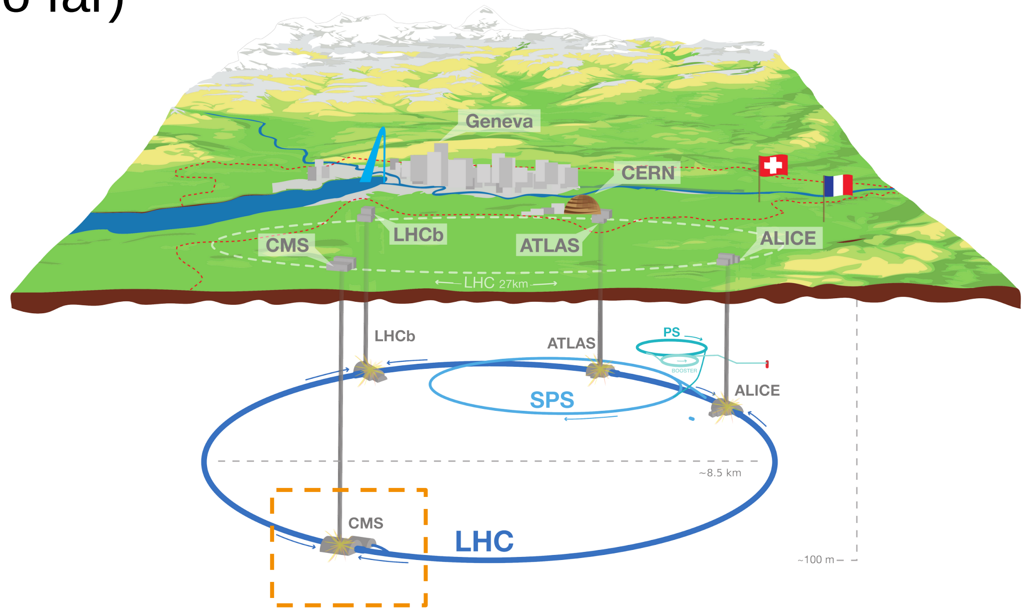


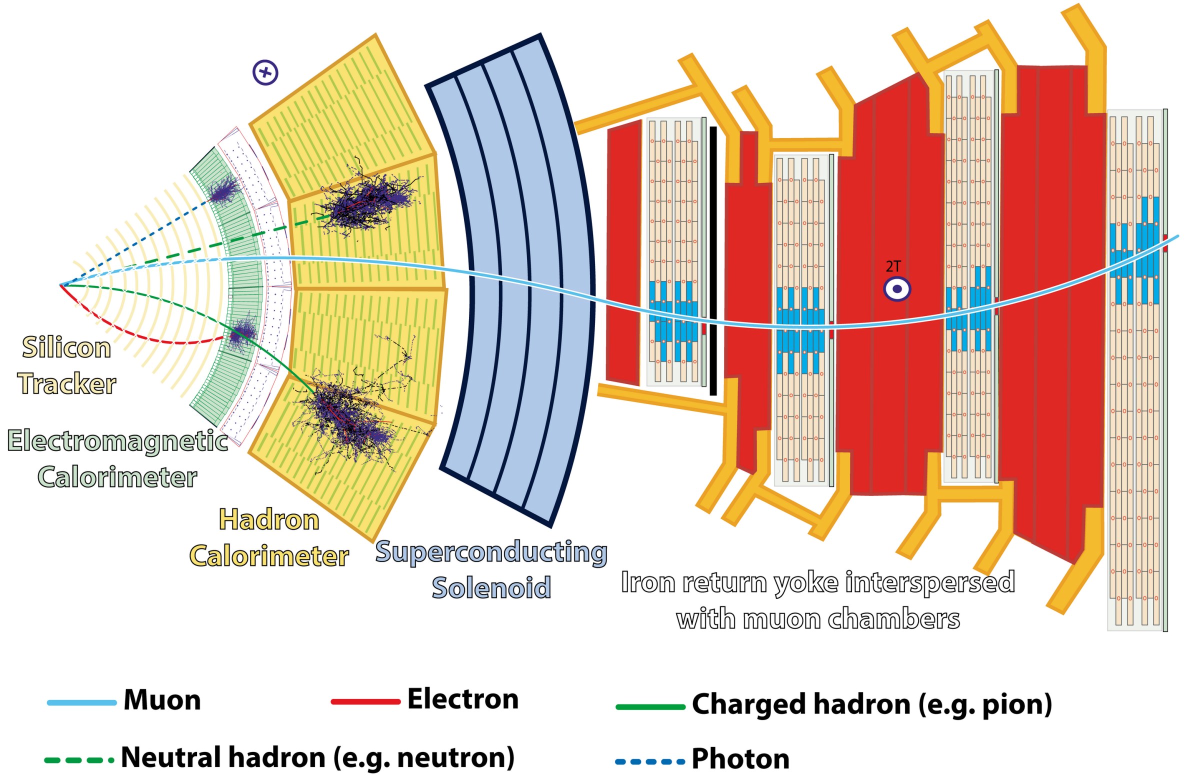
Image source: https://cds.cern.ch/record/1708849

- Four large experiments:
  ATLAS, CMS, ALICE, LHCb

The CMS Experiment

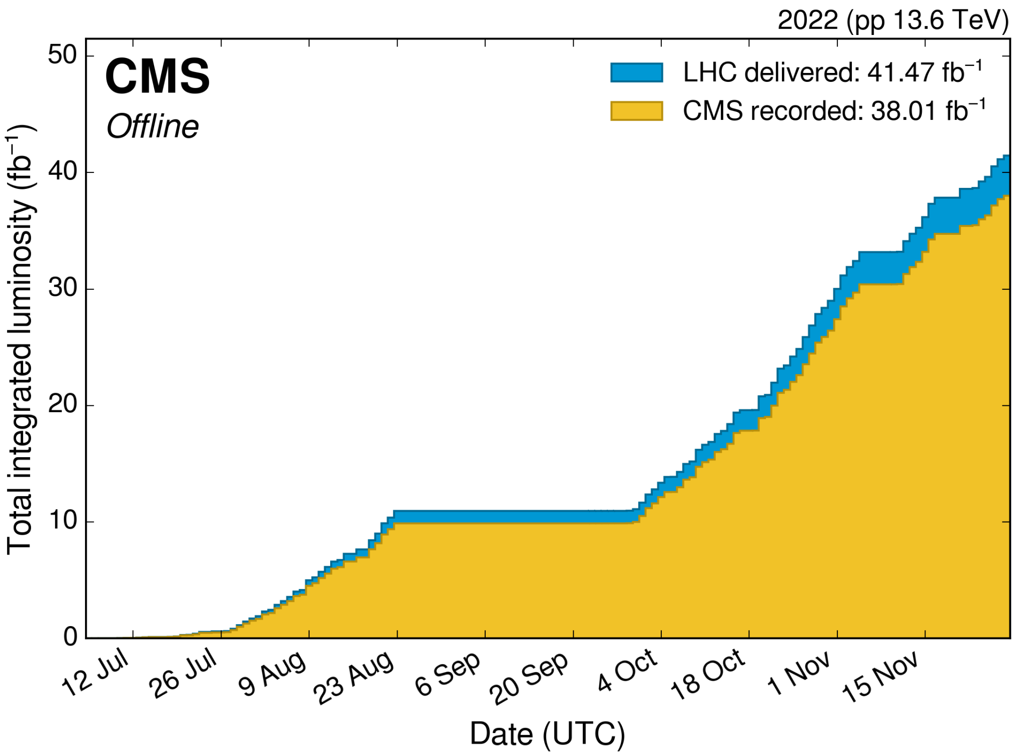
- General-purpose hermetic detector with a superconducting solenoid magnet
- Subdetectors for different particles and functions
- So far, three data-taking runs

Image source: https://cds.cern.ch/record/2120661

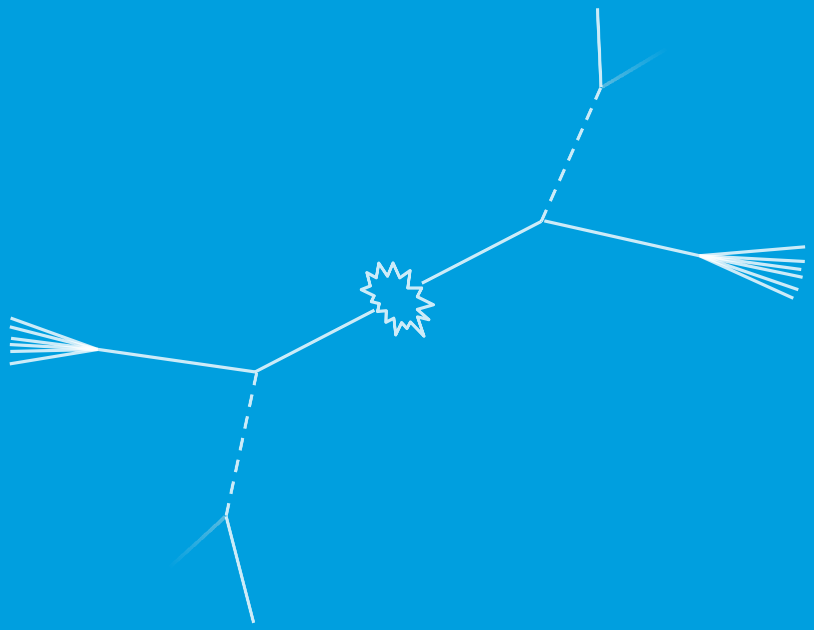


The Start of LHC Run 3

- July 2022: start of LHC Run 3
  - After three years of shutdown
  - Higher c.o.m. energy: 13.6 TeV
  - Upgraded detector components
  - Rapidly growing integrated luminosity
- Old calibrations no longer valid after long shutdown
- Need for early measurement to demonstrate good data quality



Part I

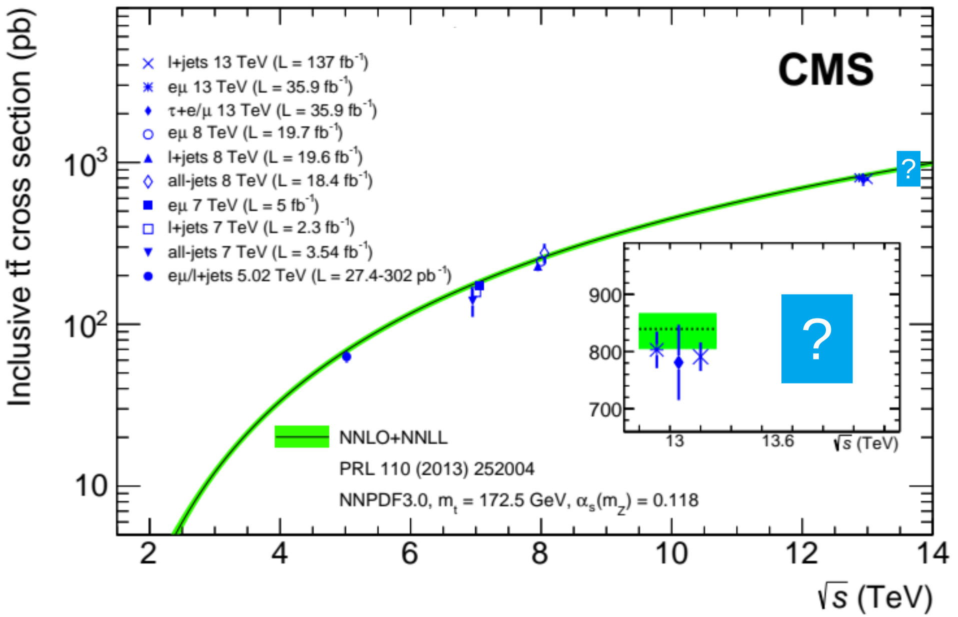


First measurement of the inclusive $t\bar{t}$ cross section in Run 3

Inclusive $t\bar{t}$ cross section

- top covers important objects at CMS: electrons, muons, (b-)jets
- Easiest top-related observable: inclusive $t\bar{t}$ production cross section
  - rises by ~10% from 13 → 13.6 TeV

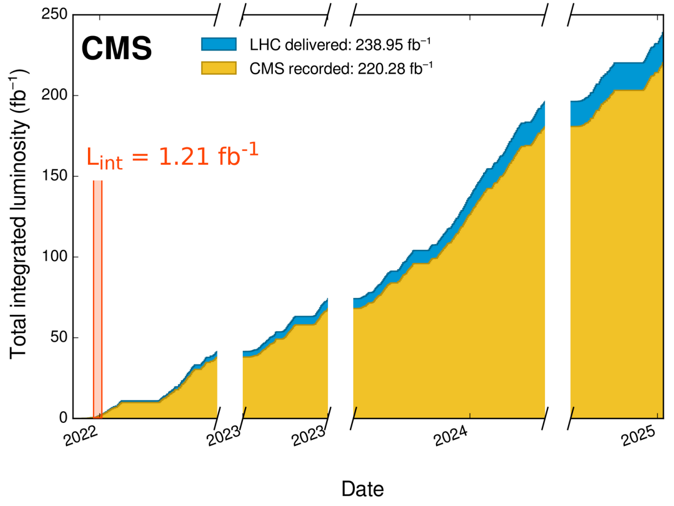
➔ Measure inclusive $t\bar{t}$ cross section in early Run 3



- Extremely short timeline:
  - Setup analysis & test on Run 2 data before Run 3 start
  - Measurement itself performed in only 2 months

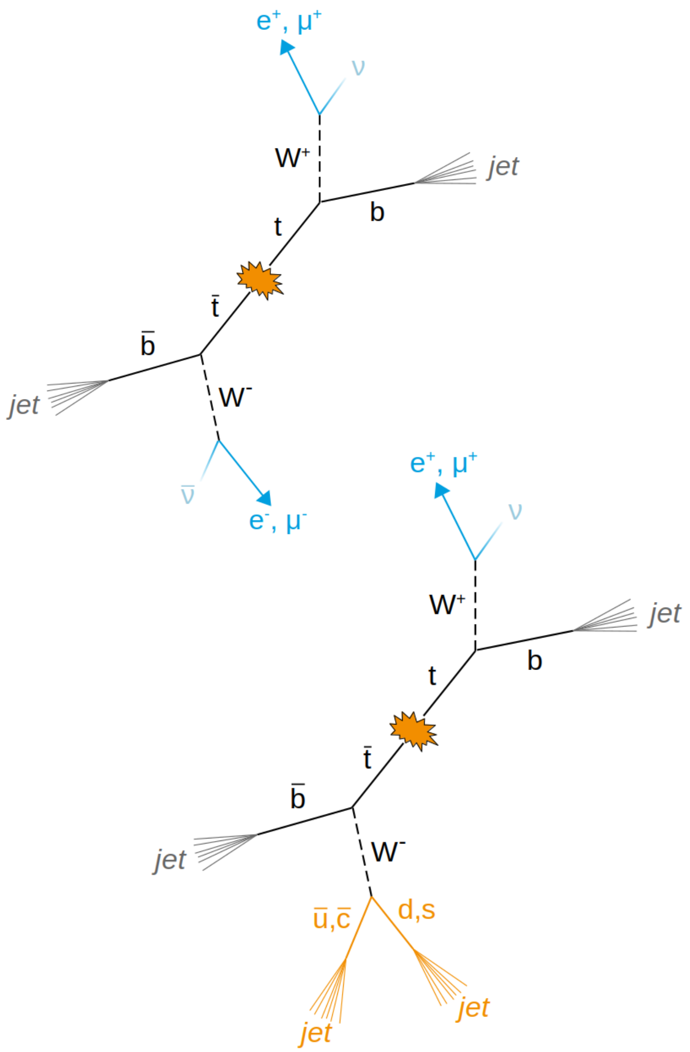
Analysis strategy

- Strategy designed from scratch, targeting early analysis
- Use ~1 week of data from Jul-Aug 2022: 1.21 fb^{-1}
- Central calibrations & corrections not yet available
  - need to estimate them or constrain them from data



Analysis strategy

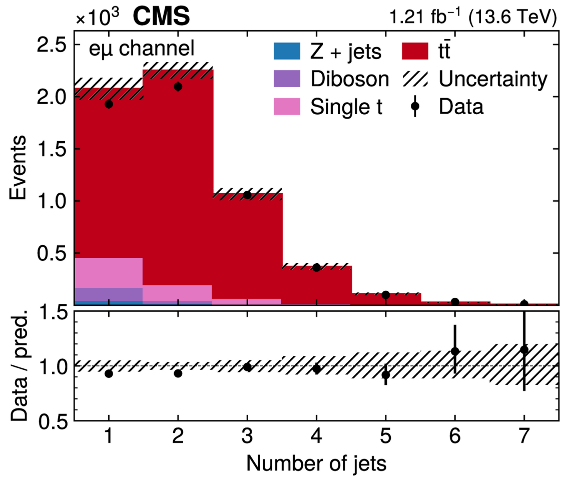
- Strategy designed from scratch, targeting early analysis
- Use ~1 week of data from Jul-Aug 2022: $1.21\ \mathrm{fb}^{-1}$
- Central calibrations & corrections not yet available
  - need to estimate them or constrain them from data
- For the first time: combine dilepton (ee, eμ, μμ) and lepton+jets (e, μ) decay channels in one likelihood fit
  - Dilepton: high purity, handle on b tagging
  - lepton+jets: high statistics
  - together: constrain lepton efficiencies



Channel definition

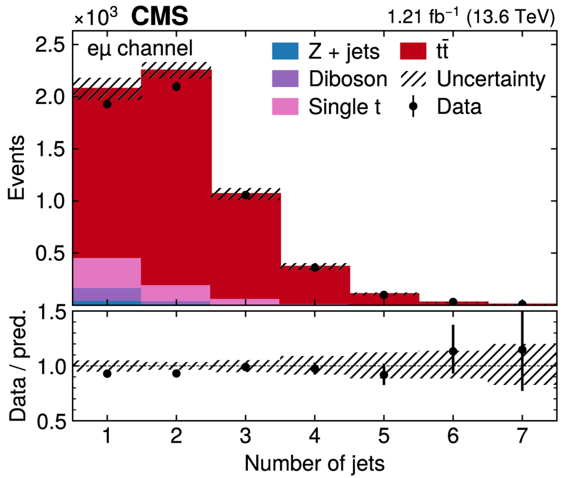
Dilepton channel: 2 opposite-sign leptons

- At least 1 jet
- For ee and μμ: large Z+jets BG
  - ➔ require at least 1 b jet
  - ➔ cut away Z peak: $|m_{\ell\ell} - m_Z| > 15$ GeV

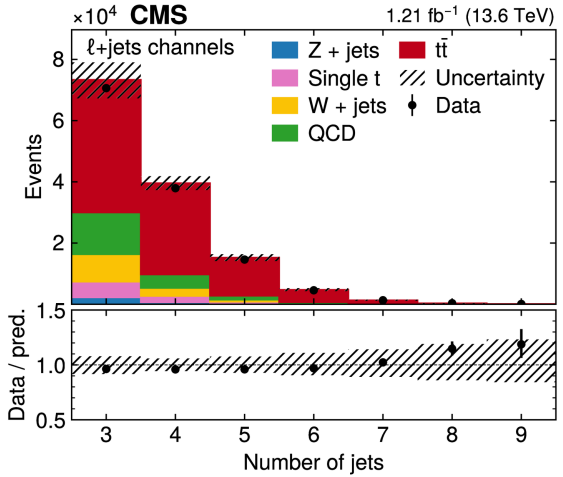


Channel definition

Dilepton channel: 2 opposite-sign leptons
- At least 1 jet
- For ee and μμ: large Z+jets BG
  - ➔ require at least 1 b jet
  - ➔ cut away Z peak: $|m_{\ell\ell} - m_Z| > 15$ GeV



Lepton+jets channel: exactly 1 lepton
- At least 3 jets
- At least 1 b tagged jet

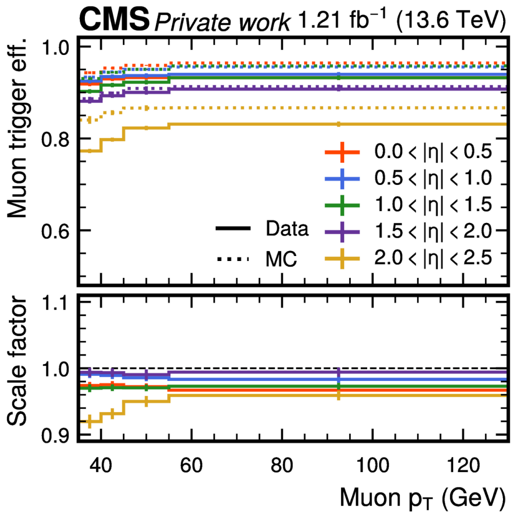


DESY.

Corrections

Most corrections derived as part of this work:

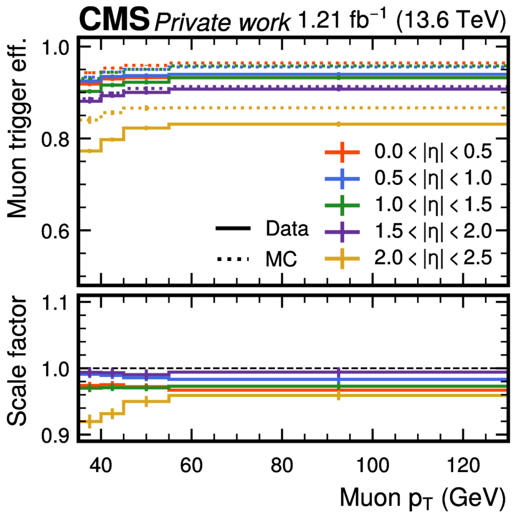
- b tagging efficiencies – in situ from data in the likelihood fit
- Trigger efficiencies – derived with tag&probe method
- Pileup corrections – data-driven reweighting to data



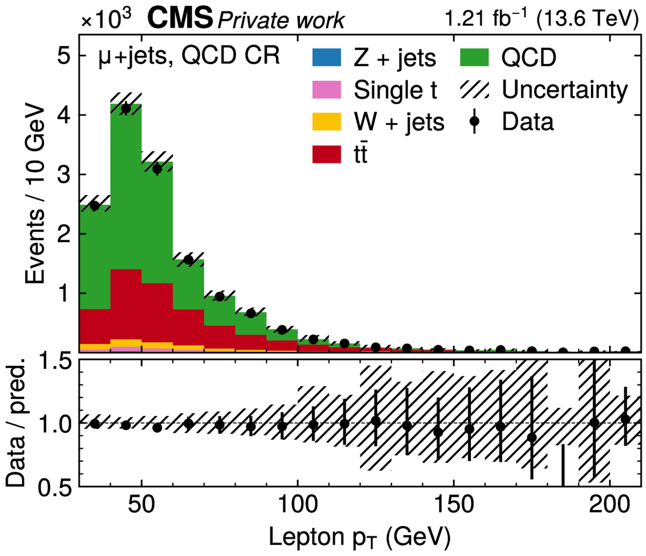
Corrections

Most corrections derived as part of this work:

- b tagging efficiencies – in situ from data in the likelihood fit
- Trigger efficiencies – derived with tag&probe method
- Pileup corrections – data-driven reweighting to data



- QCD multijet BG with fake/non-prompt lepton
  – data-driven using ABCD method
  in lepton isolation sideband



- Normalization of Z+jets background
  – from data close to Z peak

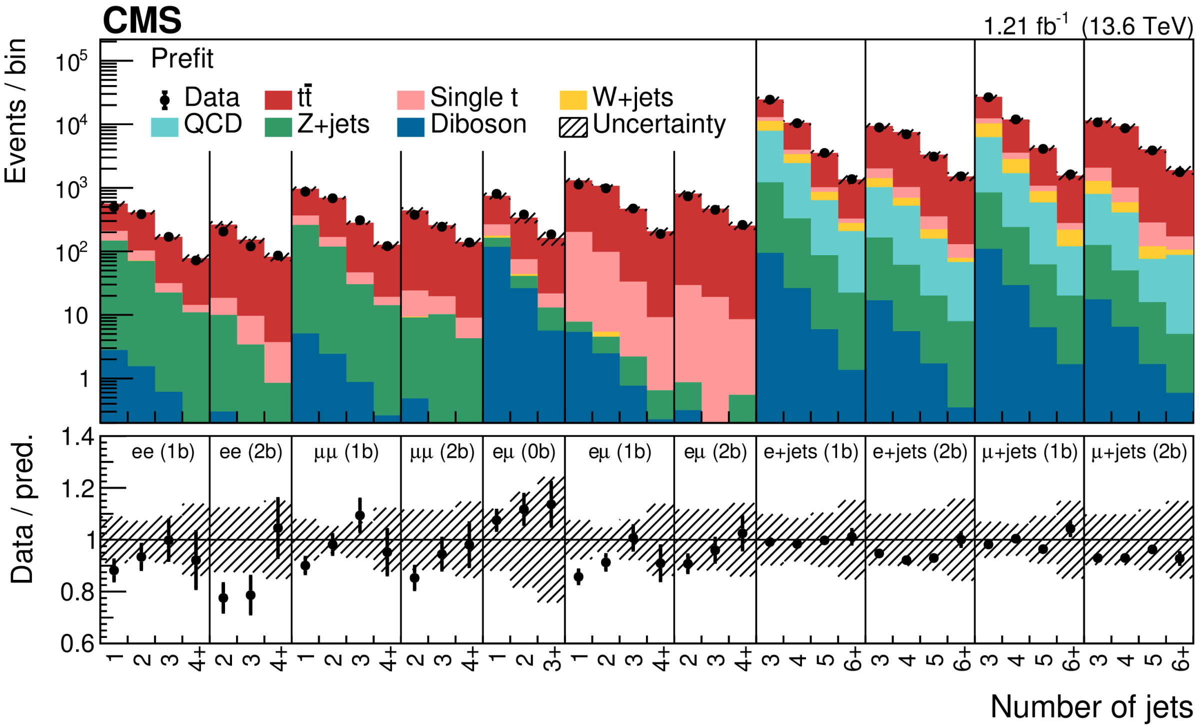
Results: prefit

- Channels defined by
  - Lepton content

→ Further separated by
  - b jet content:
    0, 1, or 2 b jets

→ Coarsely binned in
  - Number of jets

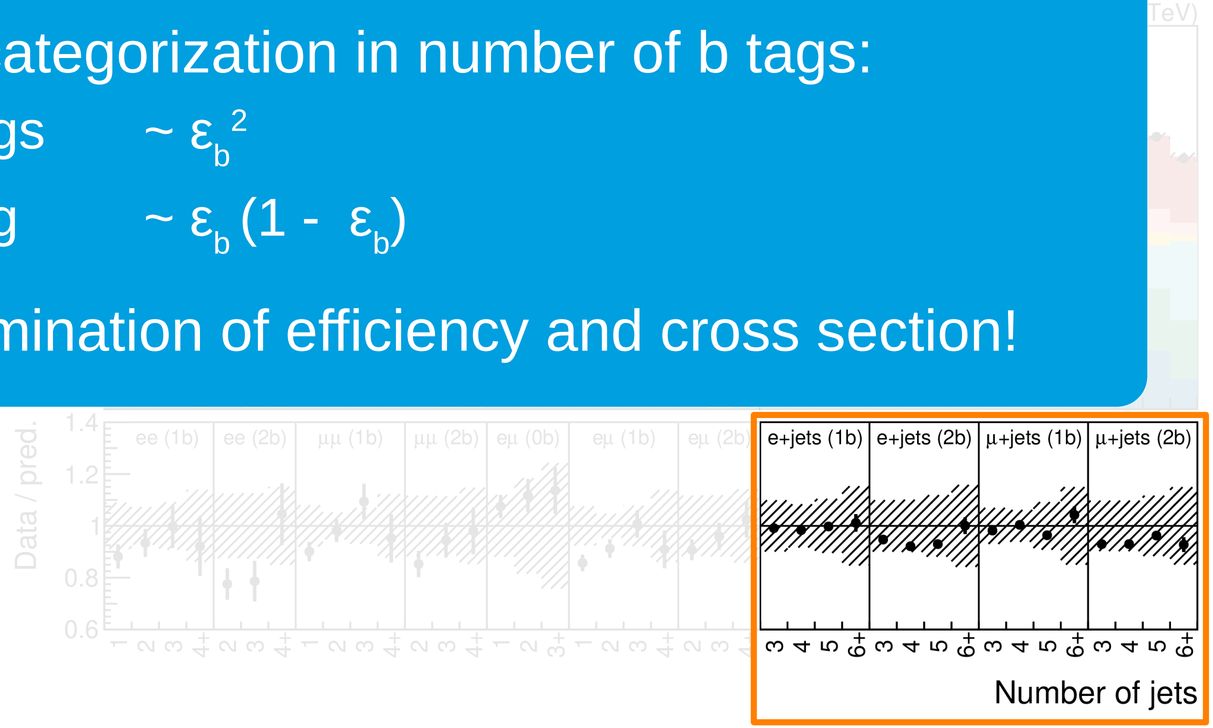
Note: no b jet SF applied, no b tagging uncertainties



Results: prefit

- Difference in b tagging efficiency ε_b in data and MC
- Sensitivity through categorization in number of b tags:
  - Events with 2 b tags $\sim \varepsilon_b^2$
  - Events with 1 b tag $\sim \varepsilon_b (1 - \varepsilon_b)$
- Simultaneous determination of efficiency and cross section!

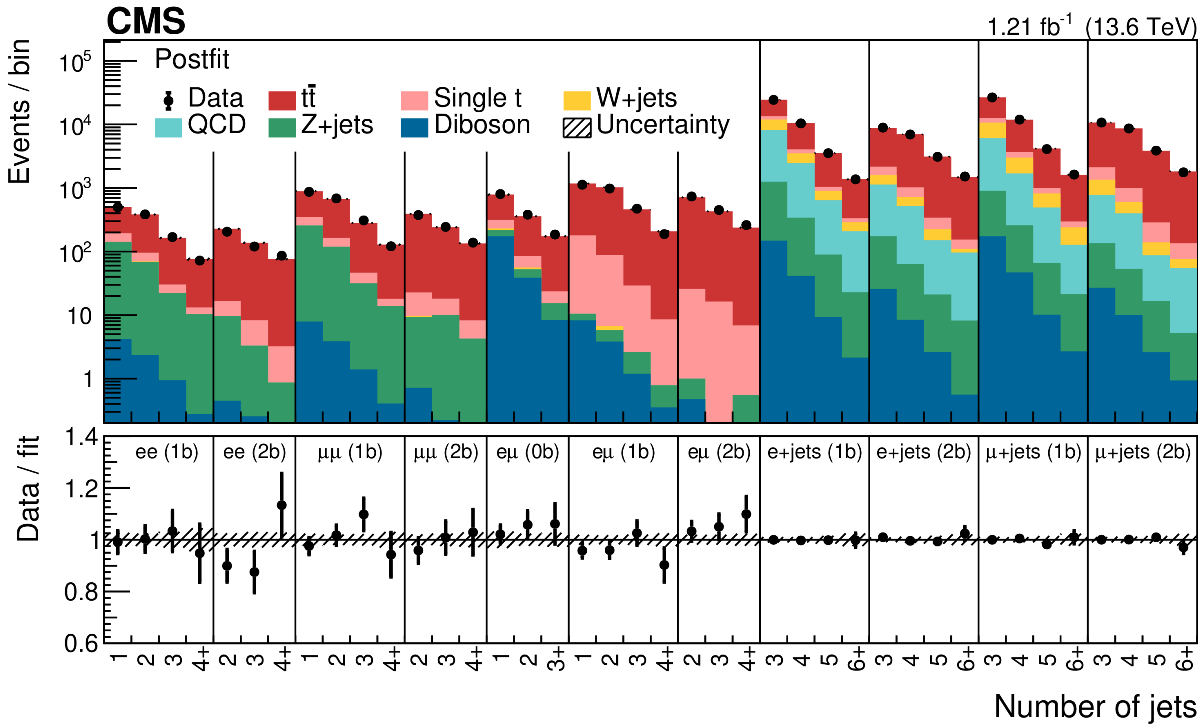
- Number of jets



Note: no b jet SF applied, no b tagging uncertainties

Results: postfit

- Profile likelihood fit with nuisance parameters
- b tag efficiency freely floating
- Postfit agreement is improved significantly!



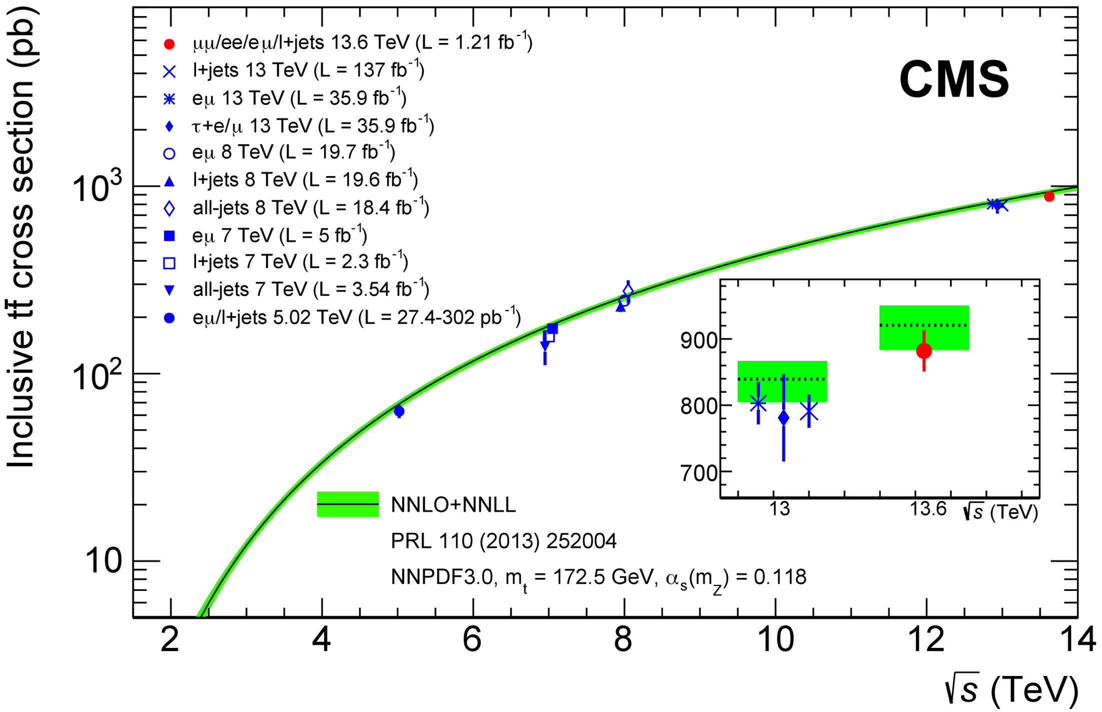
Results: cross section

JHEP 08 (2023) 204

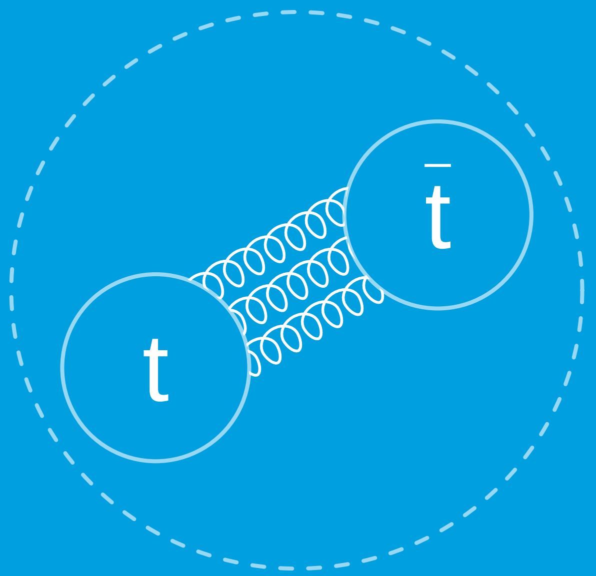
$$\sigma_{t\bar{t}} = 882 \pm 23(\mathrm{stat} + \mathrm{syst}) \pm 20(\mathrm{lumi})\mathrm{pb}$$

- ~ 3.5% total uncertainty
- Competetive with previous measurements despite early timeline
- Dominant uncertainty sources:
  - Luminosity
  - Lepton ID efficiency
  - b tag efficiency
- In agreement with theory:

$$\sigma_{t\bar{t}}^{\mathrm{pred}} = 921^{+29}_{-37}\mathrm{pb}$$



Part II

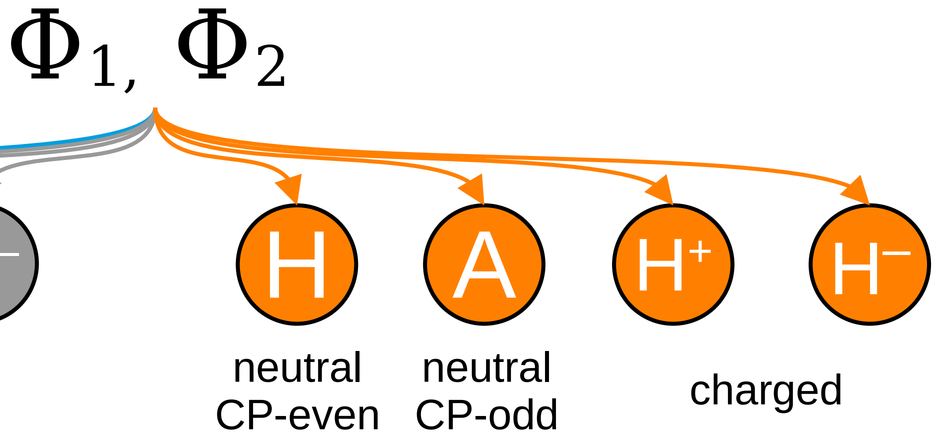


Search for (pseudo)scalar bosons in $t\bar{t}$ events and $t\bar{t}$ bound states

Additional Higgs bosons

- SM Higgs: complex SU(2) doublet ϕ → one real scalar after symmetry breaking
- Many BSM models predict additional Higgs bosons
  e.g. Two-Higgs Doublet Model (2HDM), Supersymmetry, …

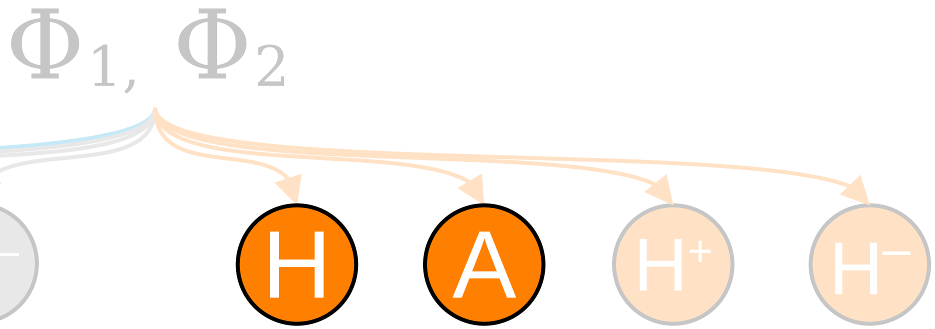
Example: 2HDM



Additional Higgs bosons

- SM Higgs: complex SU(2) doublet ϕ → one real scalar after symmetry breaking
- Many BSM models predict additional Higgs bosons
  e.g. Two-Higgs Doublet Model (2HDM), Supersymmetry, …

Example: 2HDM



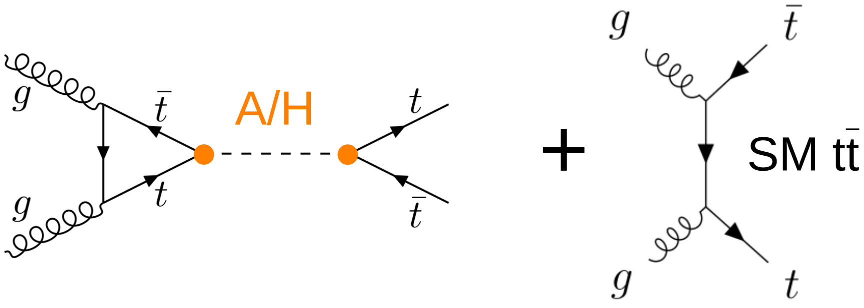
- If $m_{A/H} > 2m_t$: decay to $t\bar{t}$ → search in $t\bar{t}$ final states

Generic parameterization

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

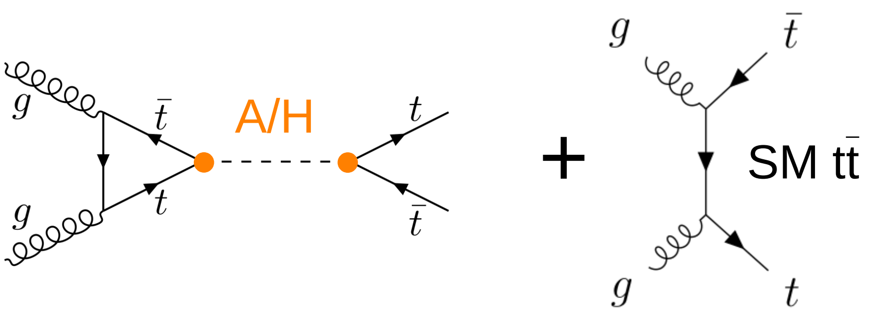
$$\mathcal{L}_H^{\text{int}} = - g_{\text{Ht}\bar{\text{t}}} \frac{m_{\text{t}}}{v} \bar{\text{t}} \text{t} \text{H}$$

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



Generic parameterization

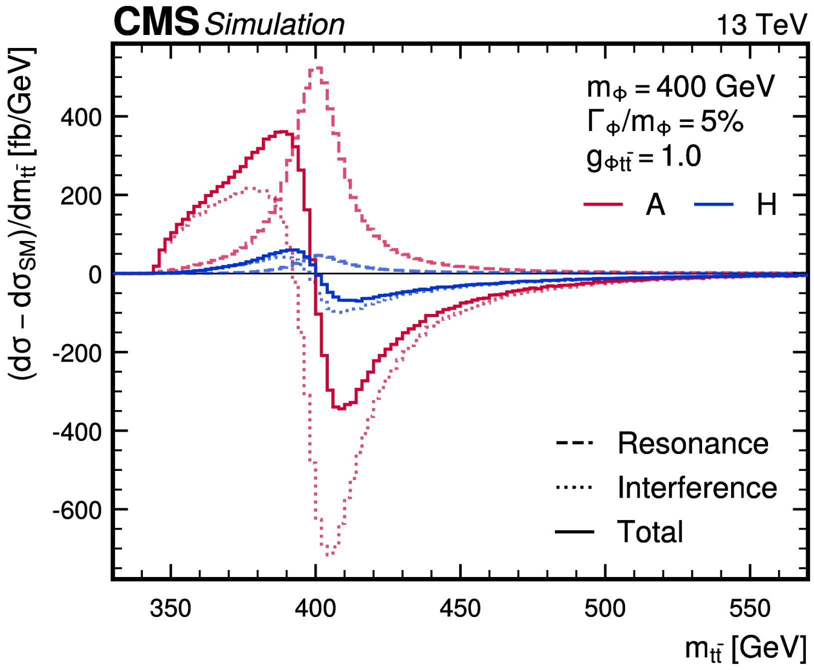
- Generic heavy pseudoscalar (A) or scalar (H) coupling solely to top quarks
- Production in gluon fusion via top quark loop



- Same final state as SM $t\bar{t}$ $\rightarrow$ interference
  $\rightarrow$ peak-dip structure in invariant mass m_{tt}

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

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

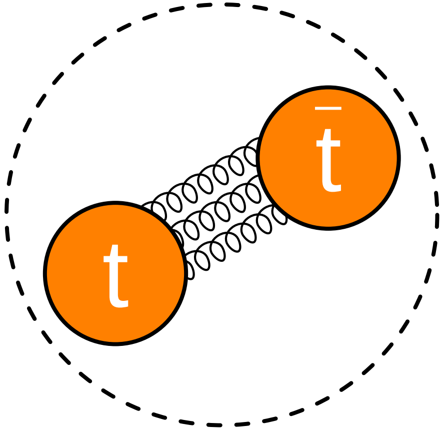


$t\bar{t}$ bound states?

- All quarks other than the top are known to form bound states (quarkonia)
- Simple estimate by analogy to hydrogen atom with QCD potential:

binding energy $E_b = -\frac{1}{2}\frac{m_\mathrm{q}}{2}\left(C^{[1]}\alpha_S\right)^2$ with $C^{[1]} = \frac{4}{3}$ for color-singlet

for $t\bar{t}$: $E_b \approx -2\,\mathrm{GeV}$ system should be bound!



$t\bar{t}$ bound states?

- All quarks other than the top are known to form bound states (quarkonia)
- Simple estimate by analogy to hydrogen atom with QCD potential:

binding energy $E_b = -\frac{1}{2}\frac{m_\mathrm{q}}{2}\left(C^{[1]}\alpha_S\right)^2$ with $C^{[1]} = \frac{4}{3}$ for color-singlet

for $t\bar{t}$: $E_b \approx -2\,\mathrm{GeV}$ system should be bound!

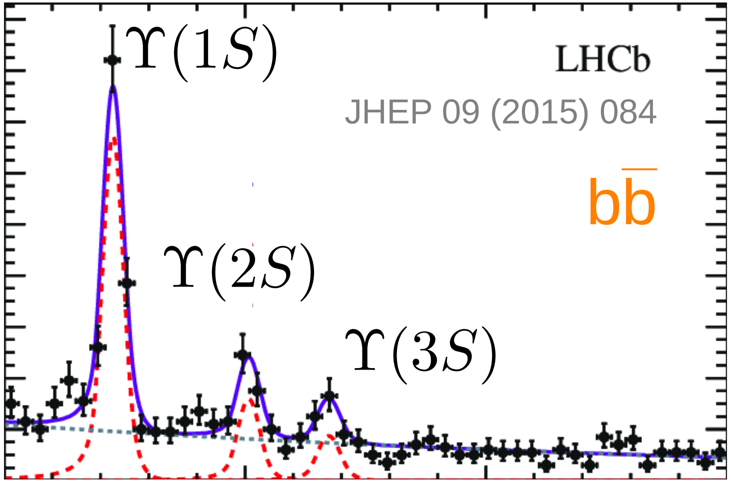
- However: lifetime of the bound state shorter than classical revolution time

$$\tau_{\mathrm{t\bar{t}}} = \frac{1}{2}\tau_\mathrm{t} = 2.4 \times 10^{-25}\,\mathrm{s} \quad < \quad \tau_\mathrm{rev} = 1 \times 10^{-24}\,\mathrm{s}$$

- Fraction of $t\bar{t}$ systems that live long enough: $\approx 1\%$ small effect!

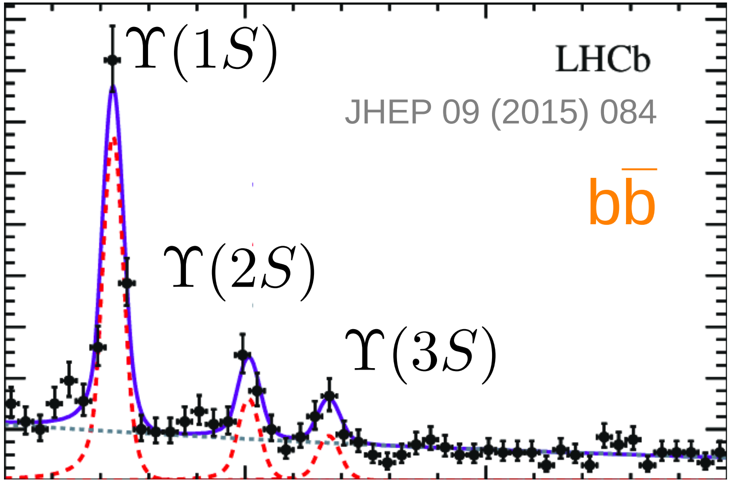
$t\bar{t}$ bound states: NRQCD

- more quantitative: non-relativistic QCD (NRQCD)
  - factorize hard scattering and long-distance effects (from Schrödinger equation)

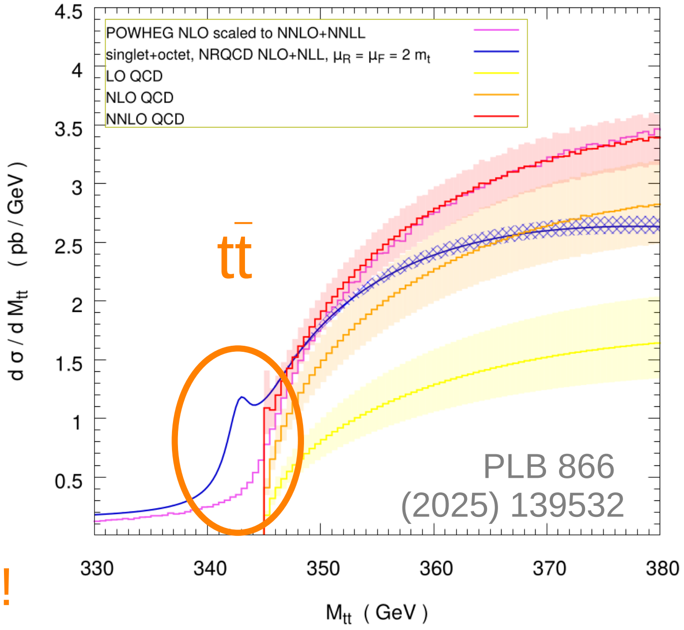


$t\bar{t}$ bound states: NRQCD

- more quantitative: non-relativistic QCD (NRQCD)
  - factorize hard scattering and long-distance effects (from Schrödinger equation)



finite width →



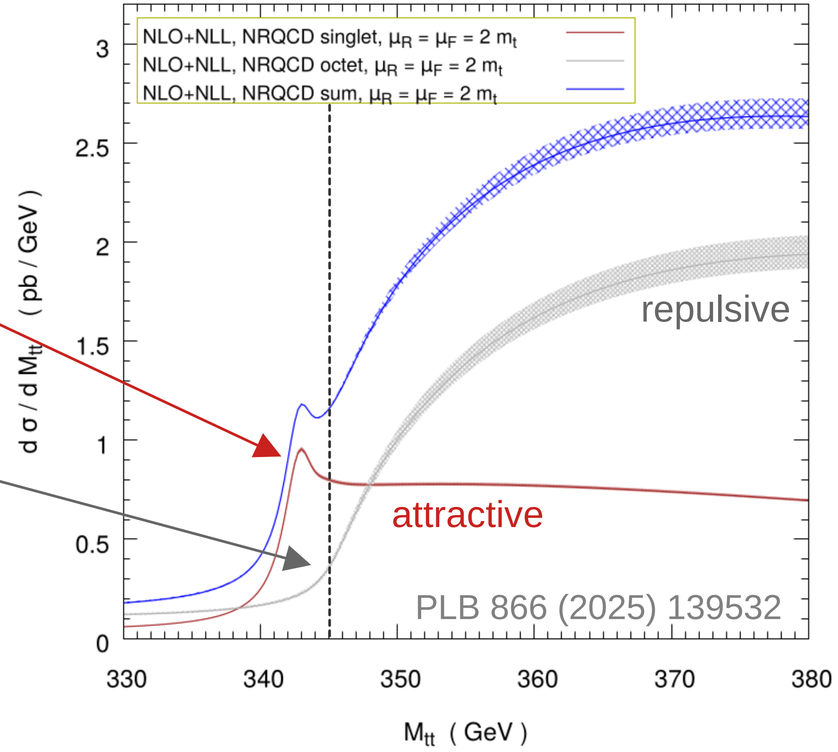
one broad peak around the $t\bar{t}$ threshold!

”quasi-bound state”

$t\bar{t}$ bound states: NRQCD

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

- Results of NRQCD calculations:
  - Color-singlet ($^1S_0^{[1]}$) - attractive
    $\rightarrow$ Peak below the $t\bar{t}$ threshold
    CP-odd / pseudoscalar spin state!
  - Color-octet ($^1S_0^{[8]}$ or $^3S_1^{[8]}$) - repulsive
    $\rightarrow$ Suppressed below the $t\bar{t}$ threshold



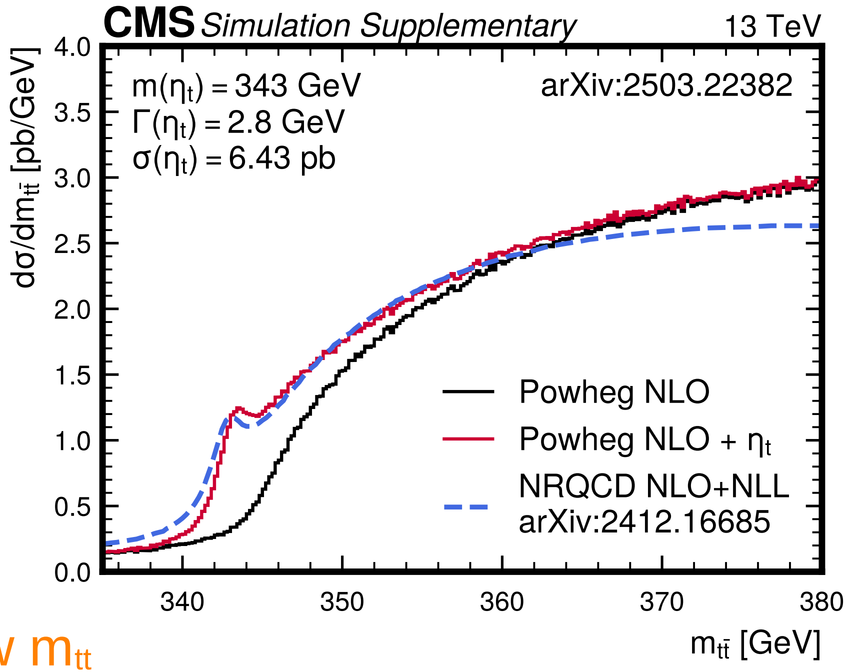
$t\bar{t}$ bound states: modeling in MC

- Cannot use NRQCD to produce events for analysis...
- Use simplified model for MC simulation: η_t (PRD 104 (2021) 034023, JHEP 03 (2024) 099)
  - Generic spin-0, color-singlet state η_t
  - Couplings to gluons and tops (pseudoscalar)
  - Fit mass and width from NRQCD:

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

$$\Gamma_{\eta_t} \approx 2\Gamma_t = 2.8\,\mathrm{GeV}$$

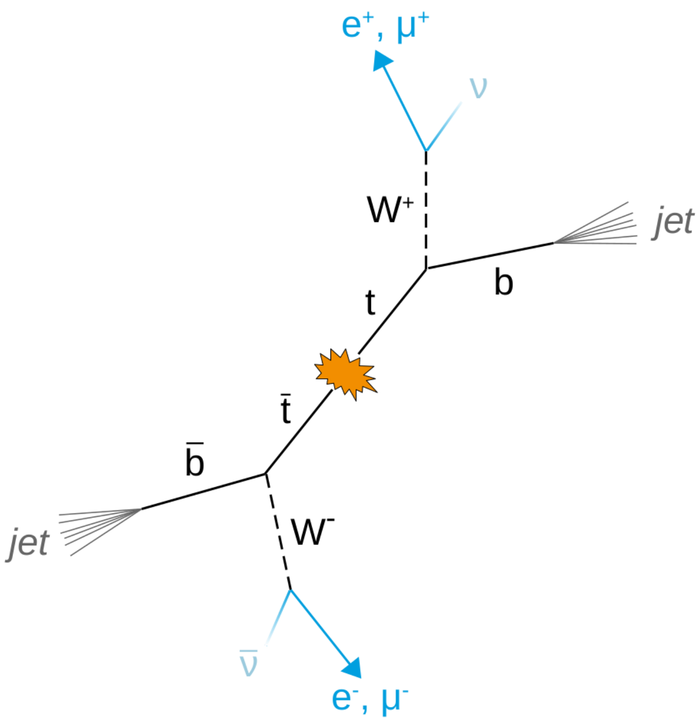
➔ Stack on top of ordinary pQCD $t\bar{t}$ simulation!
Fits well to NRQCD prediction



- Result: pseudoscalar-like enhancement at low $m_{t\bar{t}}$

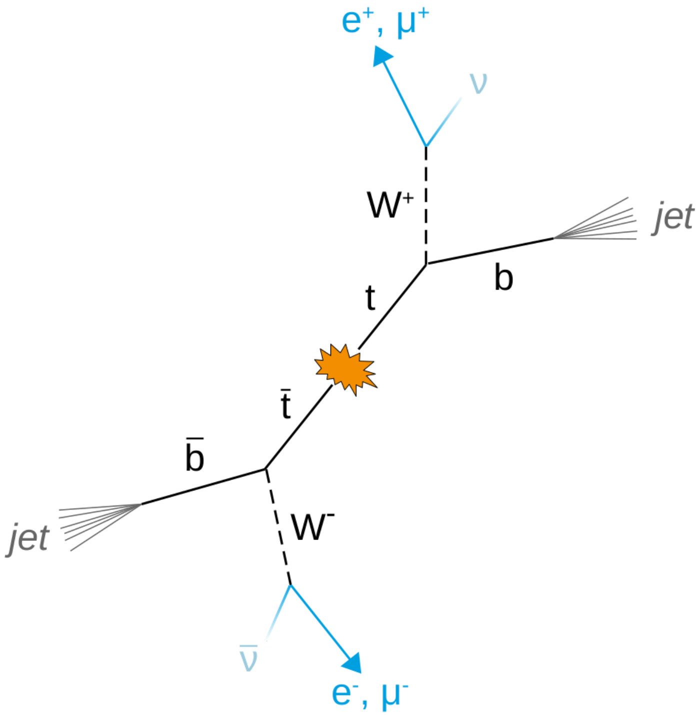
Analysis strategy

- Consider dilepton final state of $t\bar{t}$
  - Full Run 2 dataset: 138 fb^{-1}
  - Selection: 2 leptons, ≥2 jets, ≥1 b tag
    in ee/μμ: reject Z peak, require p_T^{miss} > 40 GeV



Analysis strategy

- Consider dilepton final state of $t\bar{t}$
  - Full Run 2 dataset: 138 fb^{-1}
  - Selection: 2 leptons, ≥2 jets, ≥1 b tag
    in ee/μμ: reject Z peak, require p_T^{miss} > 40 GeV
- Analytic reconstruction of $t\bar{t}$ system:
  - Assign b jets using likelihood based on $m_{\ell b}$
  - Assumptions: all p_T^{miss} from νν, tops and Ws on-shell
  - Solve fourth-order polynomial equations
  - Finite detector resolution: repeat reconstruction 100 times
    with randomly smeared inputs, take weighted average
- Four-momenta of top and antitop
  → invariant $t\bar{t}$ mass, and...

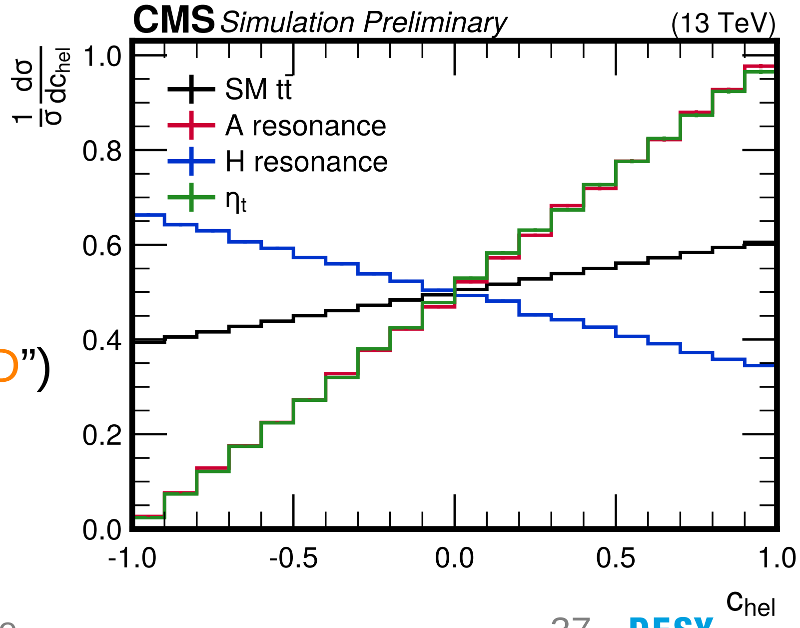


Spin correlation observables

- 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

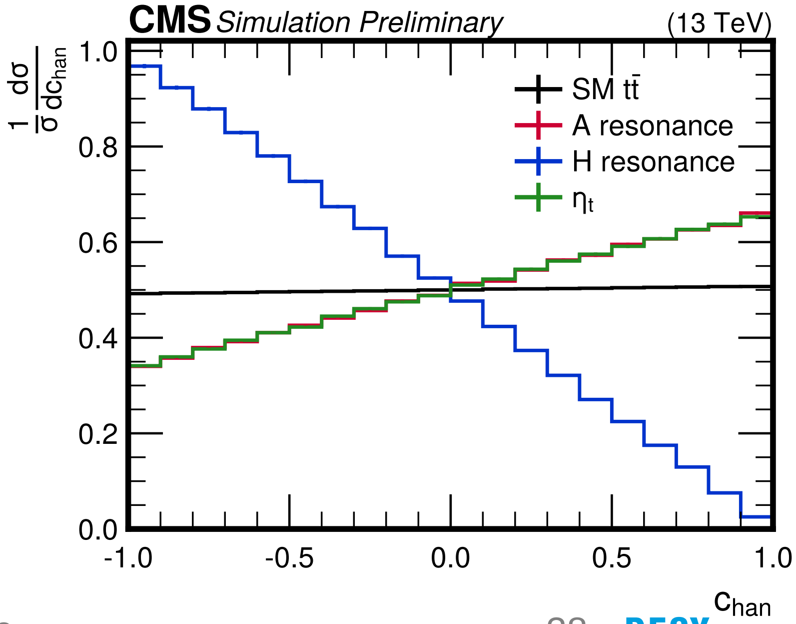
Spin correlation observables

- 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 → transfer spin information to decay products
- Construct spin correlation observables from tops & leptons

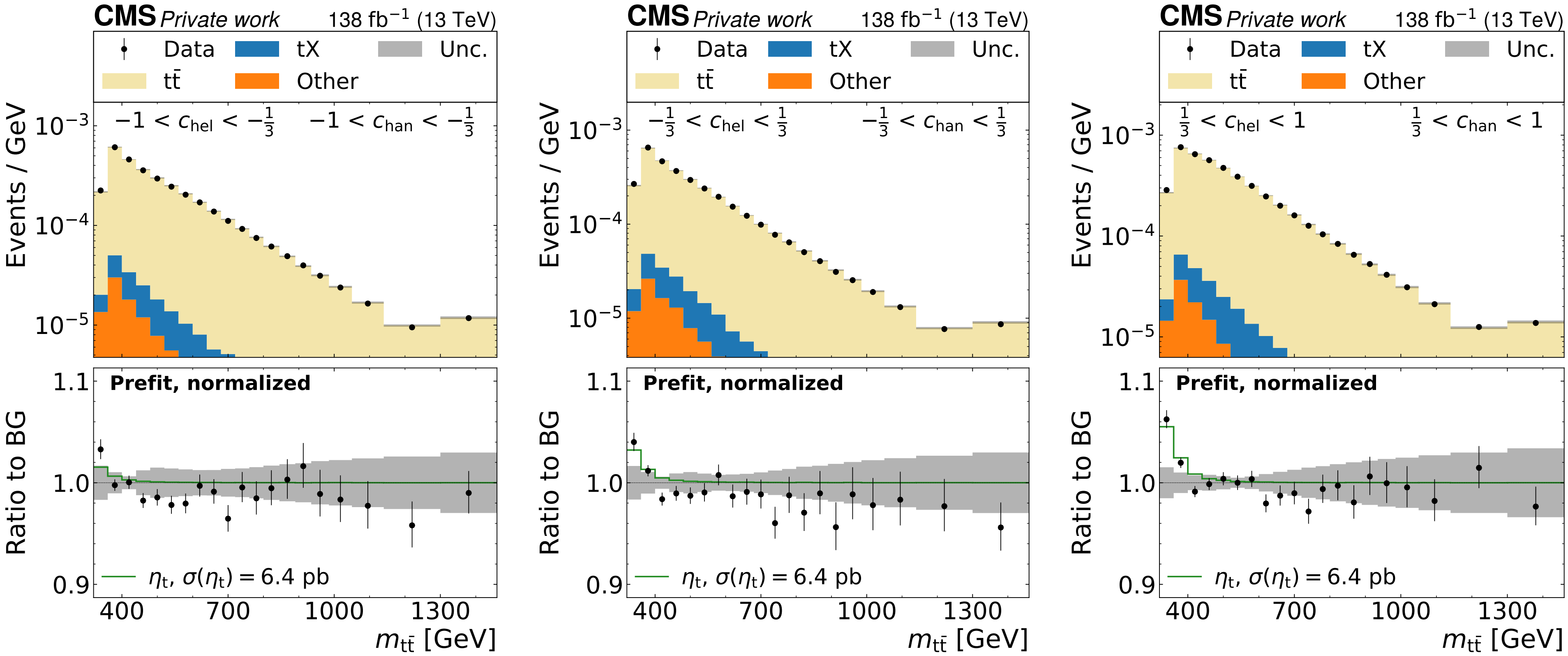


- Variable #2: c_{han}
  - Similar as c_{hel}, separating scalars from SM
  - Maximally negative slope for $^3P_0$ state
  - Construct similarly from lepton momenta, with sign flip for component parallel to top momentum

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

Prefit distributions

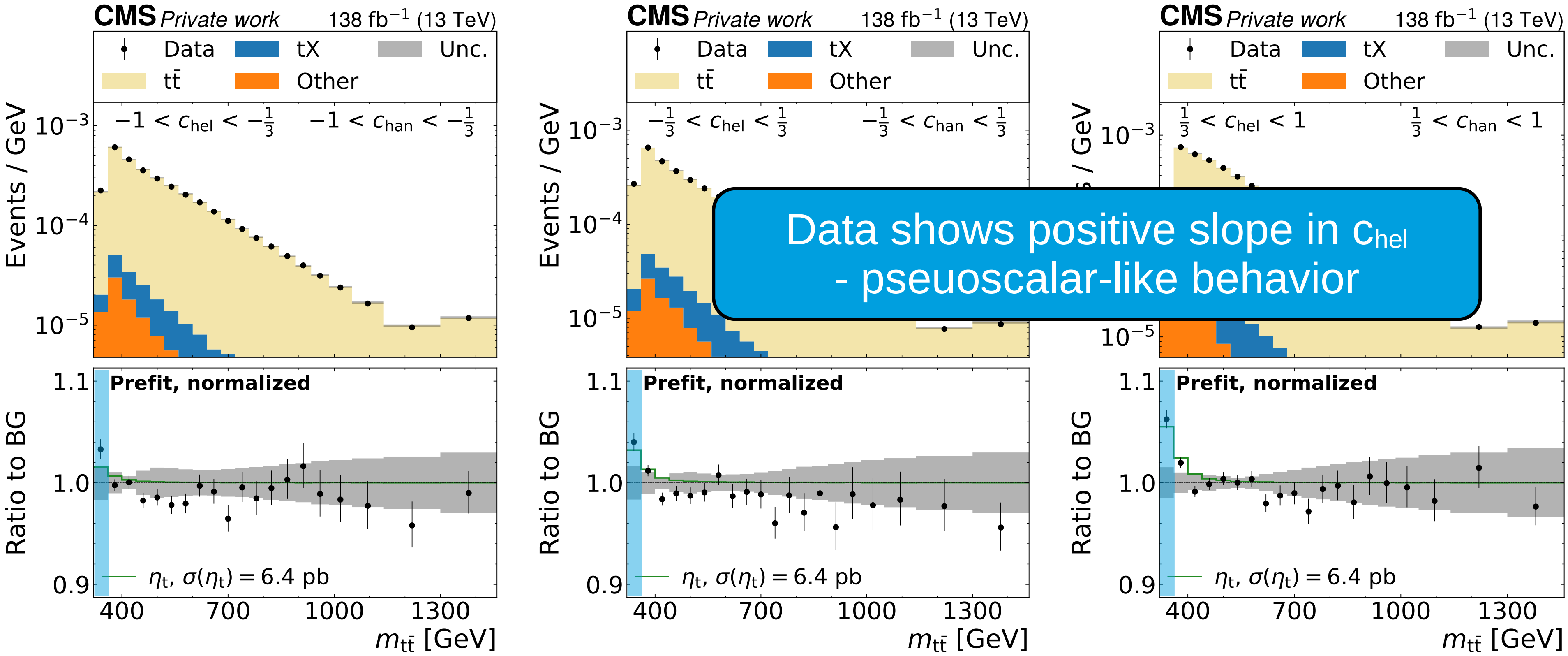
Differences between data and prediction observed in low m_{tt} bins!



DESY.

Prefit distributions

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



Data shows positive slope in c_{hel}
- pseuoscalar-like behavior

Cross section measurement

- Extract cross section using the simplified η_t color-singlet model
- "cross section" = difference to perturbative prediction

Cross section measurement

- Extract cross section using the simplified η_t color-singlet model
- "cross section" = difference to perturbative prediction

$$\sigma(\eta_\mathrm{t}) = 8.7 \pm 0.5(\mathrm{stat}) \pm 1.0(\mathrm{syst})\,\mathrm{pb} = 8.7 \pm 1.1\,\mathrm{pb}$$

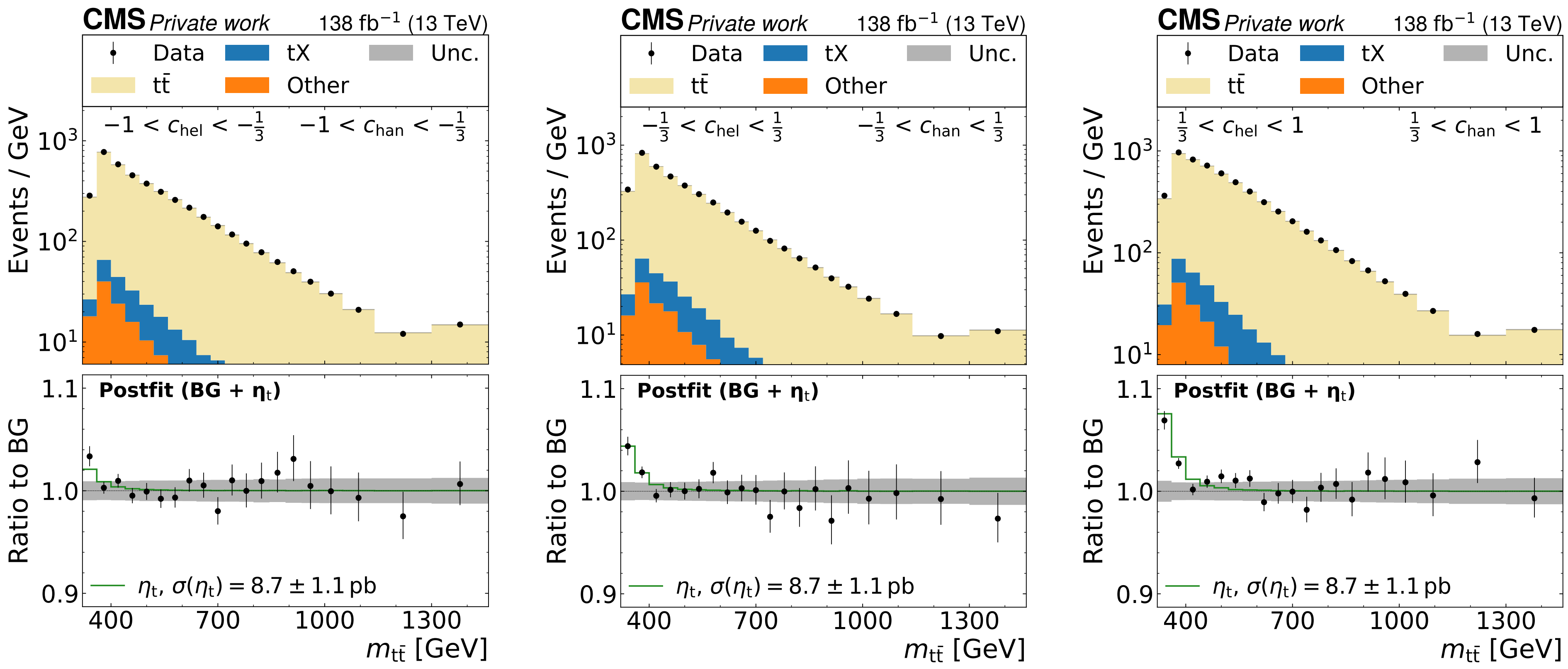
> 5σ significance!

- Main uncertainties: $t\bar{t}$ background modeling (PRD 104 (2021) 034023)
- Same order of magnitude as NRQCD estimate: $\sigma(\eta_t)^{\mathrm{pred}} = 6.4\,\mathrm{pb}$
- Confirmed by ATLAS in preliminary result (ATLAS-CONF-2025-008)

published as RoPP 88 087801 (2025)

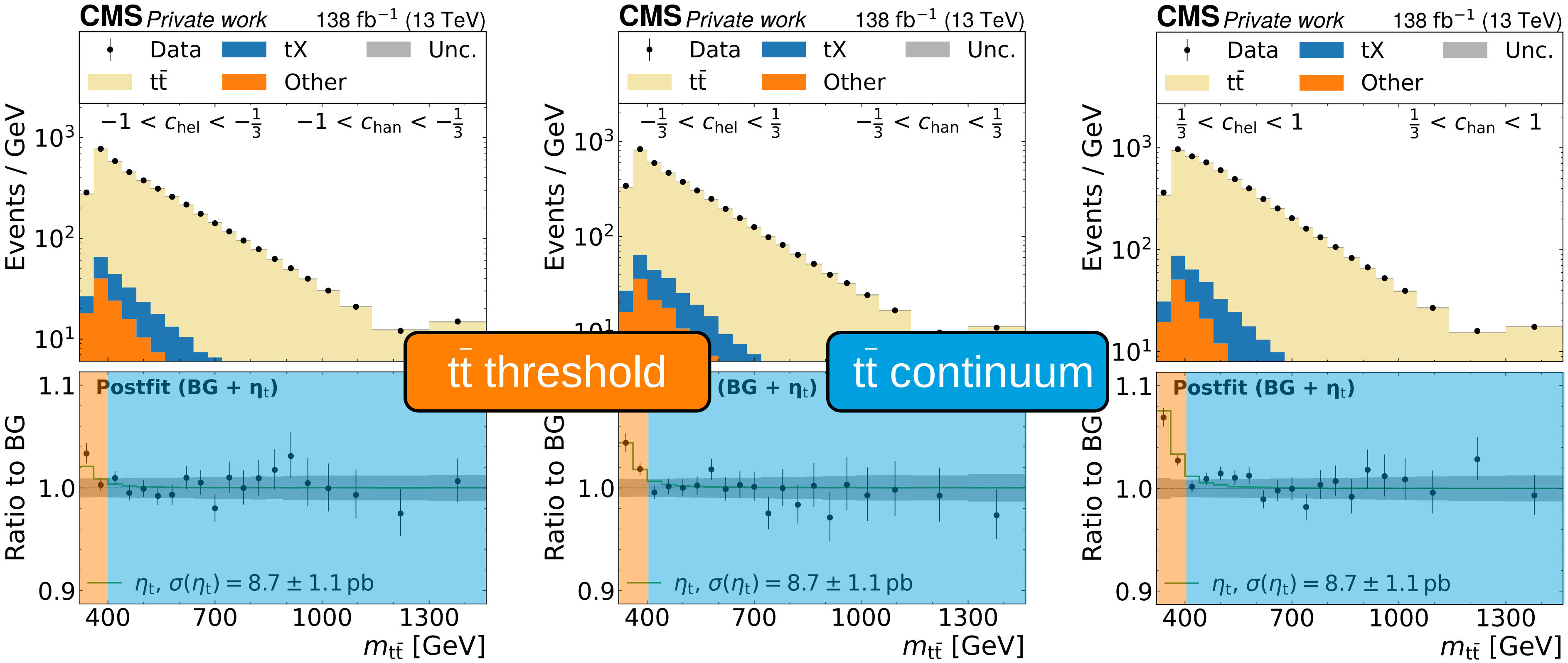
Postfit distributions: η_t

η_t model describes the data well after the fit

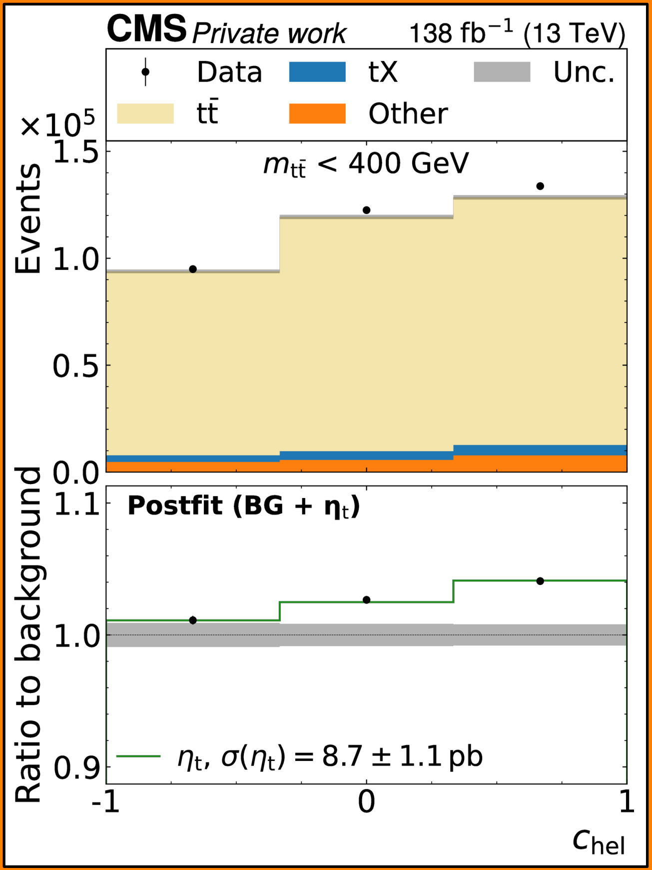


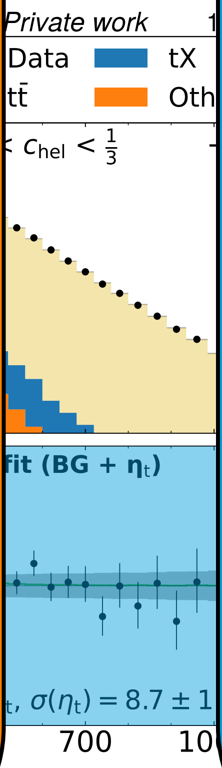
Postfit distributions: η_t

η_t model describes the data well after the fit

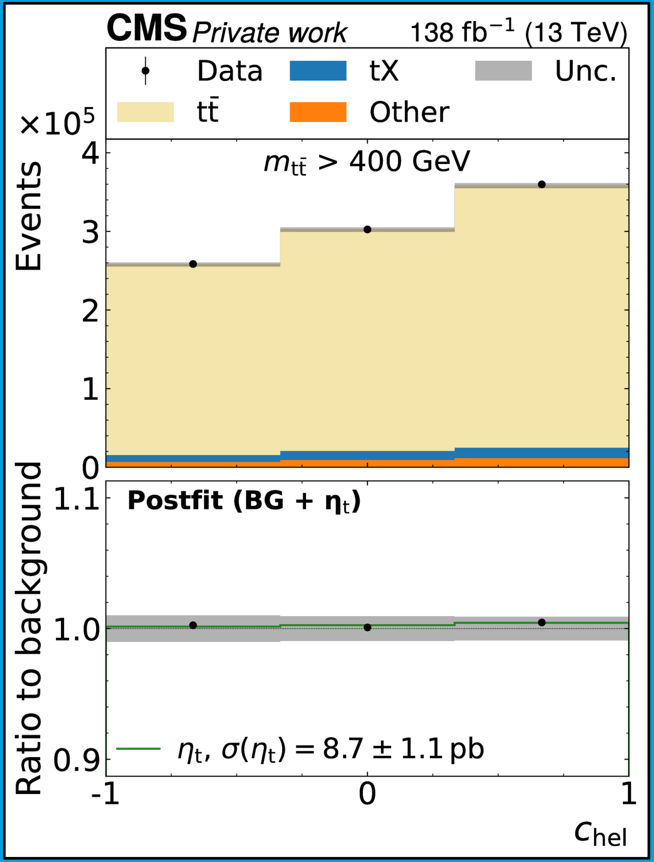


$t\bar{t}$ threshold: Data shows slope in c_{hel}



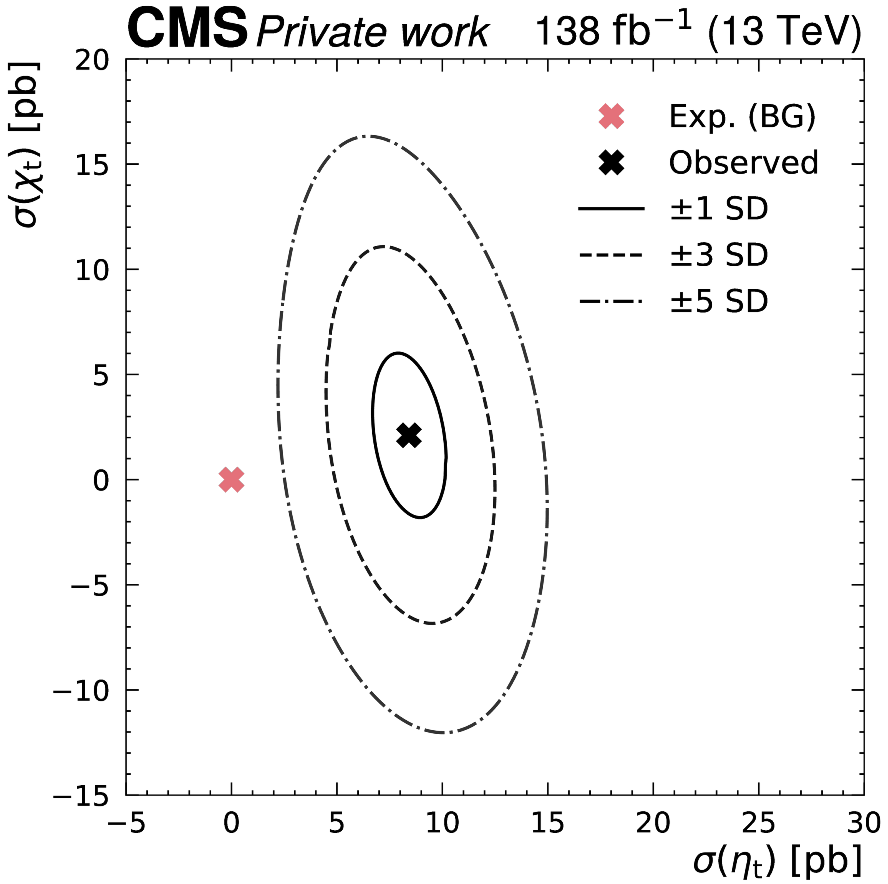


$t\bar{t}$ continuum: No slope in data



Scalar vs. pseudoscalar

- Can we quantify whether the excess is scalar or pseudoscalar?
- introduce χ_t: scalar $t\bar{t}$ bound state ($^3P_0$ $t\bar{t}$ spin state)
  - similar simplified model as η_t
- Perform 2D fit with η_t and χ_t as signals
- non-zero η_t by > 5 standard deviations
- χ_t compatible with zero by 1 sigma

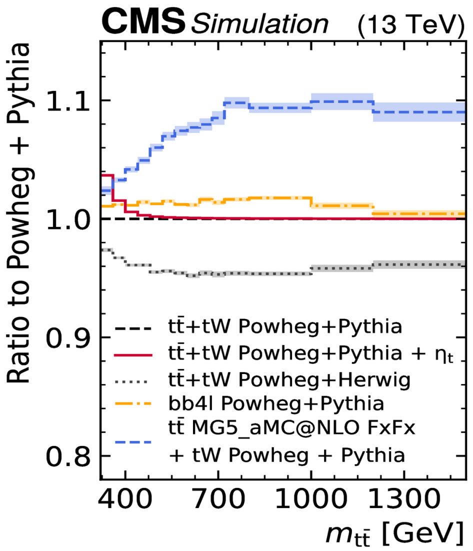


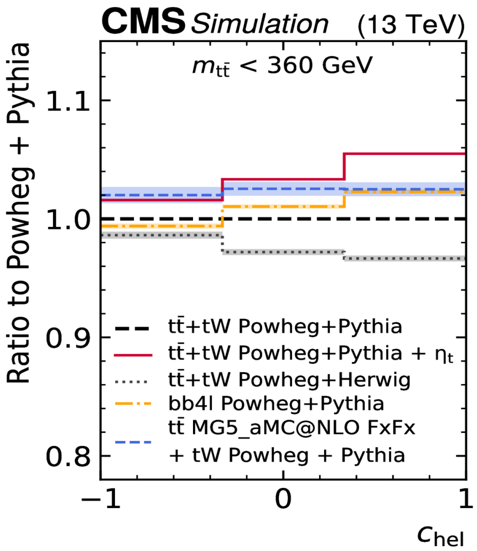
Checks of the result

How do we make sure the excess is real?

Check alternative generators for continuum $t\bar{t}$ + tW

nominal: Powheg + Pythia with NNLO QCD/NLO EW reweighting





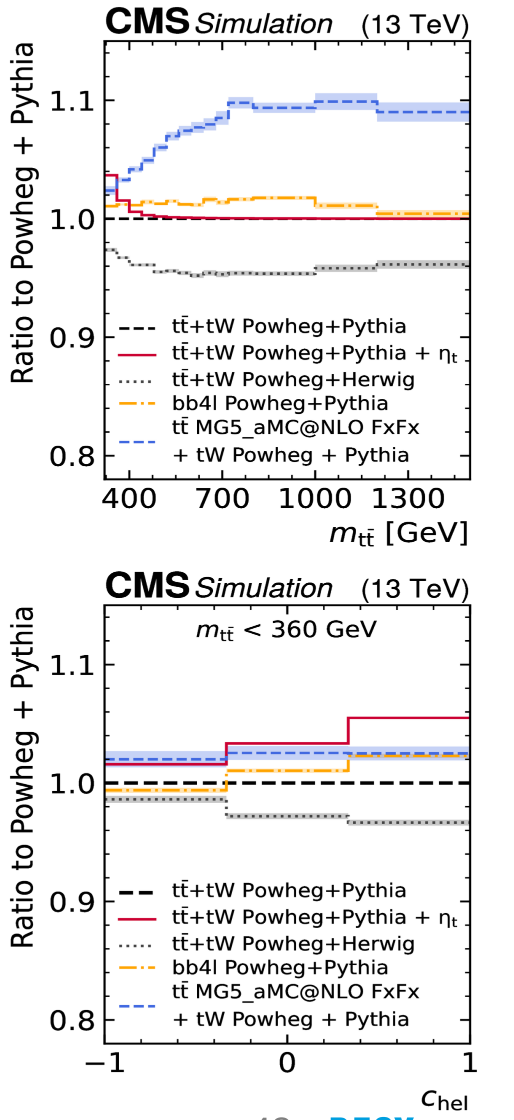
Checks of the result

How do we make sure the excess is real?

Check alternative generators for continuum $t\bar{t}$ + tW
nominal: Powheg + Pythia with NNLO QCD/NLO EW reweighting

- bb4l: full off-shell pp → bbℓℓνν at NLO in QCD, including $t\bar{t}$/tW interference
  - validated & implemented myself in CMS (service work)
  - Enhanced slope in c_{hel} w.r.t nominal Powheg $t\bar{t}$ + tW - similar to η_t
  - Reduces extracted η_t cross section:

bb4l		nominal Powheg $t\bar{t}$ + tW
$\sigma(\eta_t) = 6.6 \pm 1.4\,\text{pb}$	$\longleftrightarrow$	$\sigma(\eta_t) = 8.7 \pm 1.1\,\text{pb}$

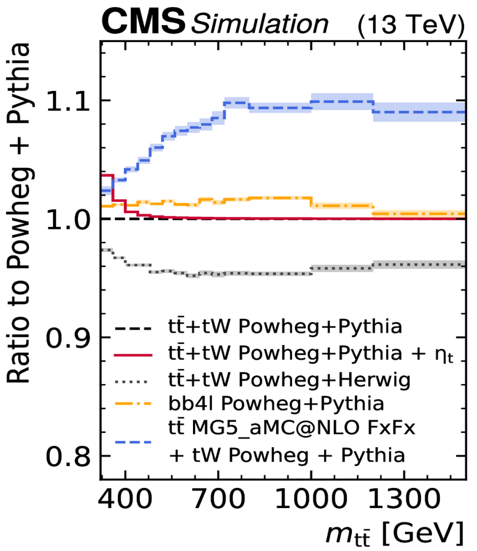


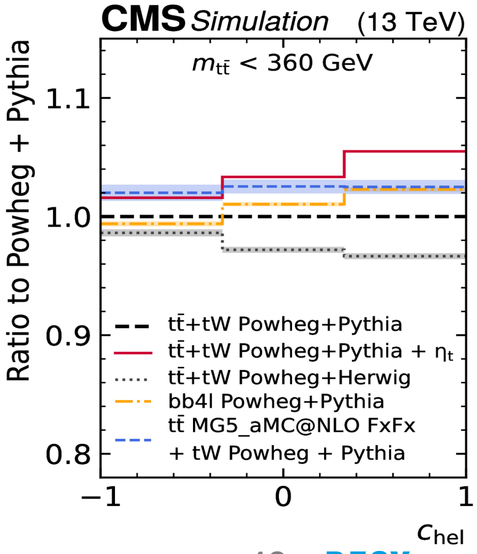
Checks of the result

How do we make sure the excess is real?

Check alternative generators for continuum $t\bar{t}$ + tW
nominal: Powheg + Pythia with NNLO QCD/NLO EW reweighting

- Parton shower for $t\bar{t}$:
  Powheg+Herwig instead of Powheg+Pythia
  - Herwig predicts more events at $t\bar{t}$ threshold
    but less spin correlation → distinguishable from η_t
  - no large change in cross section



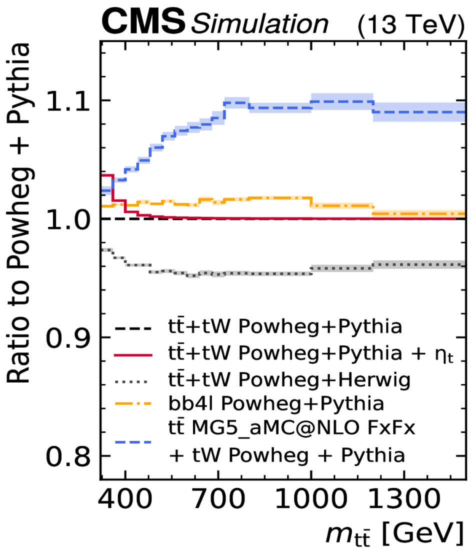


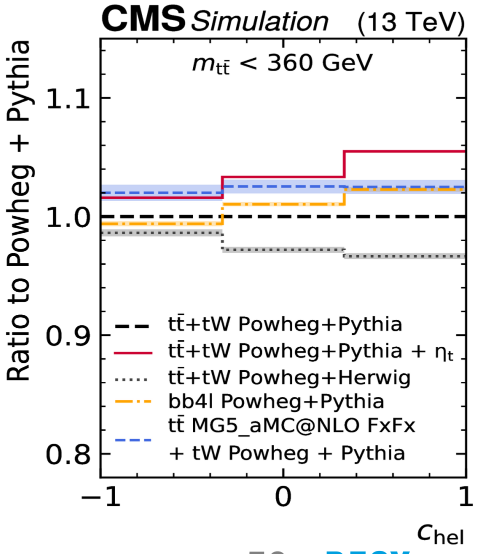
Checks of the result

How do we make sure the excess is real?

Check alternative generators for continuum $t\bar{t}$ + tW
nominal: Powheg + Pythia with NNLO QCD/NLO EW reweighting

- Parton shower for $t\bar{t}$:
  Powheg+Herwig instead of Powheg+Pythia
  - Herwig predicts more events at $t\bar{t}$ threshold but less spin correlation → distinguishable from η_t
  - no large change in cross section
- Include bb4l and Herwig as additional systematics:
  $\sigma(\eta_t) = 8.8^{+1.2}_{-1.4}\,\mathrm{pb}$ slight increase in uncertainty



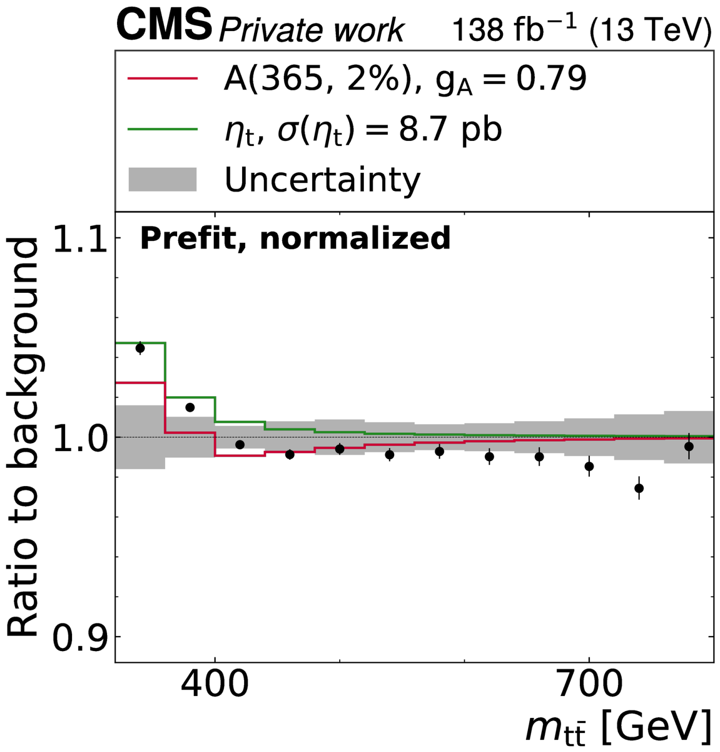


$t\bar{t}$ bound state or BSM?

- same excess can be interpreted as (BSM) pseudoscalar A: best fit

$$m_{\mathrm{A}} = 365\,\mathrm{GeV}, \quad \Gamma_{\mathrm{A}}/m_{\mathrm{A}} = 2\%, \quad g_{\mathrm{At\bar{t}}} = 0.79$$

lowest mass point probed in the analysis!



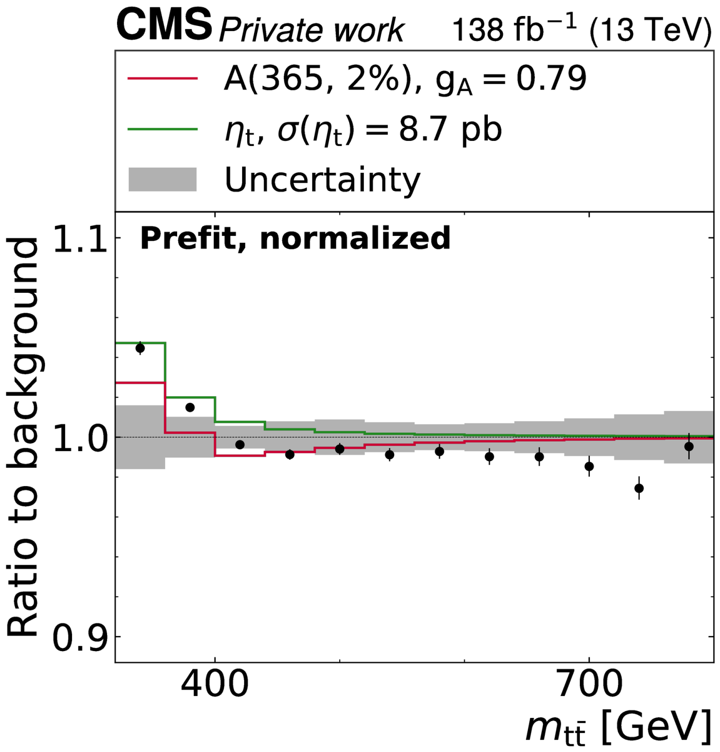
$t\bar{t}$ bound state or BSM?

- same excess can be interpreted as (BSM) pseudoscalar A: best fit

$$m_A = 365\,\text{GeV}, \quad \Gamma_A/m_A = 2\%, \quad g_{At\bar{t}} = 0.79$$

lowest mass point probed in the analysis!

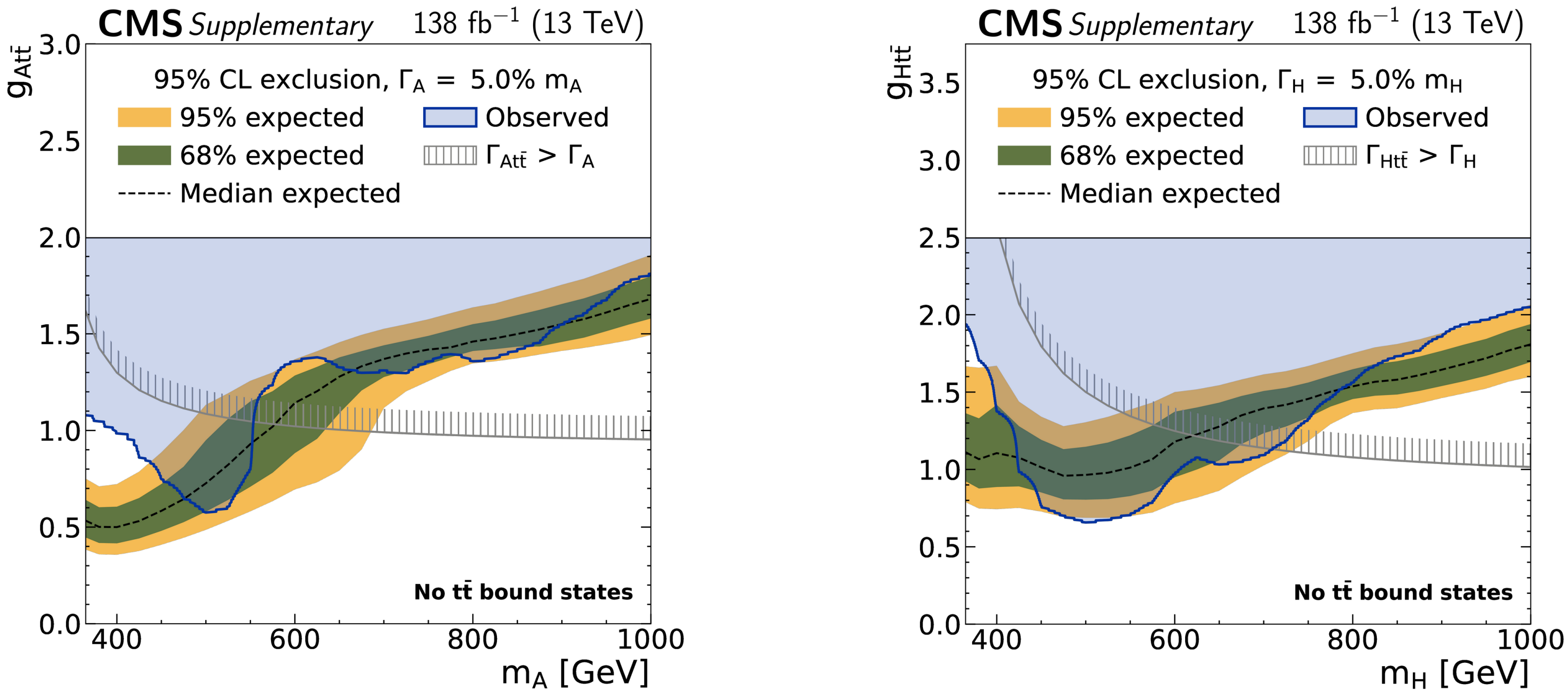
- in general: could be any combination from bound state effects and BSM
- fit slightly prefers η_t bound state, but only by 1σ



Limits on A/H

arXiv:2507.05119
submitted to *RoPP*

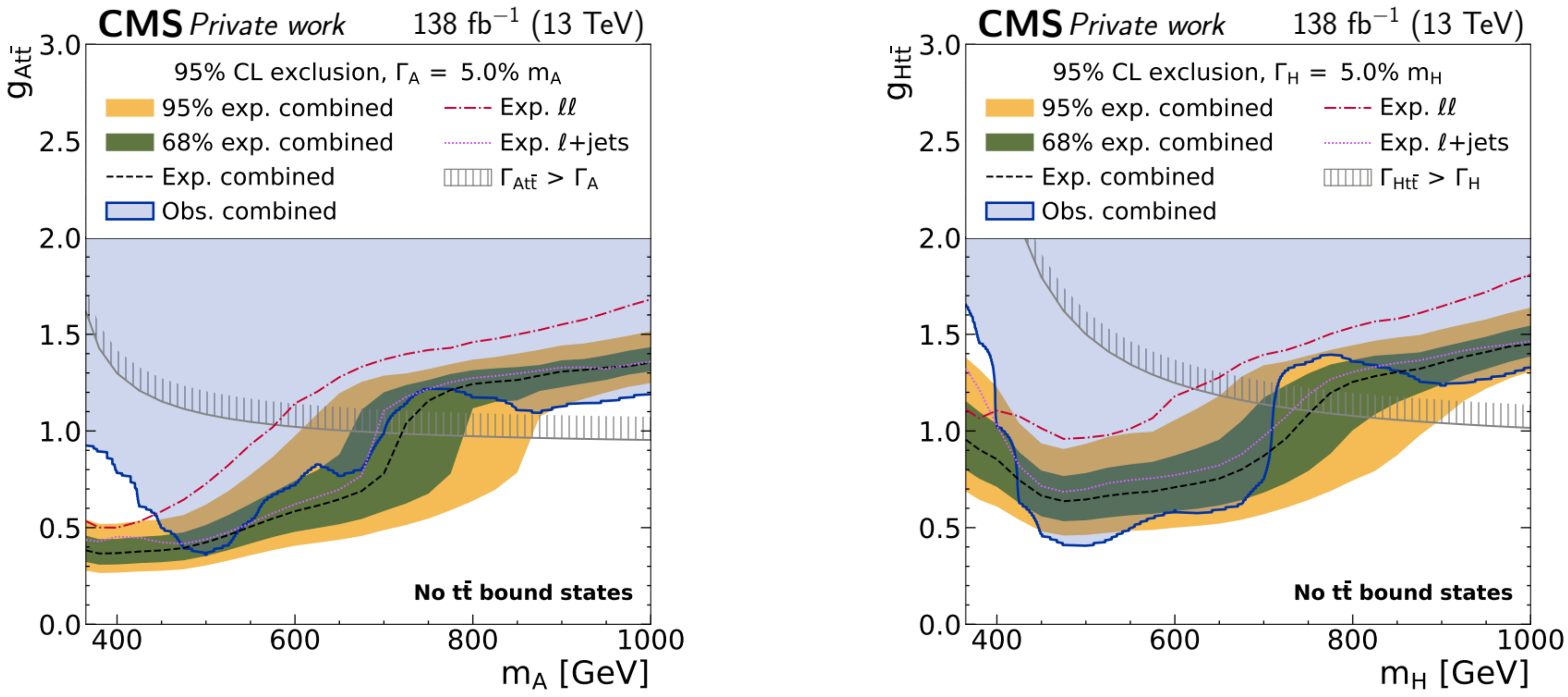
- Set limits on A/H-top coupling over large mass range
- No bound state effects considered → excess shown in limits



arXiv:2507.05119
submitted to *RoPP*

Limits on A/H: combination

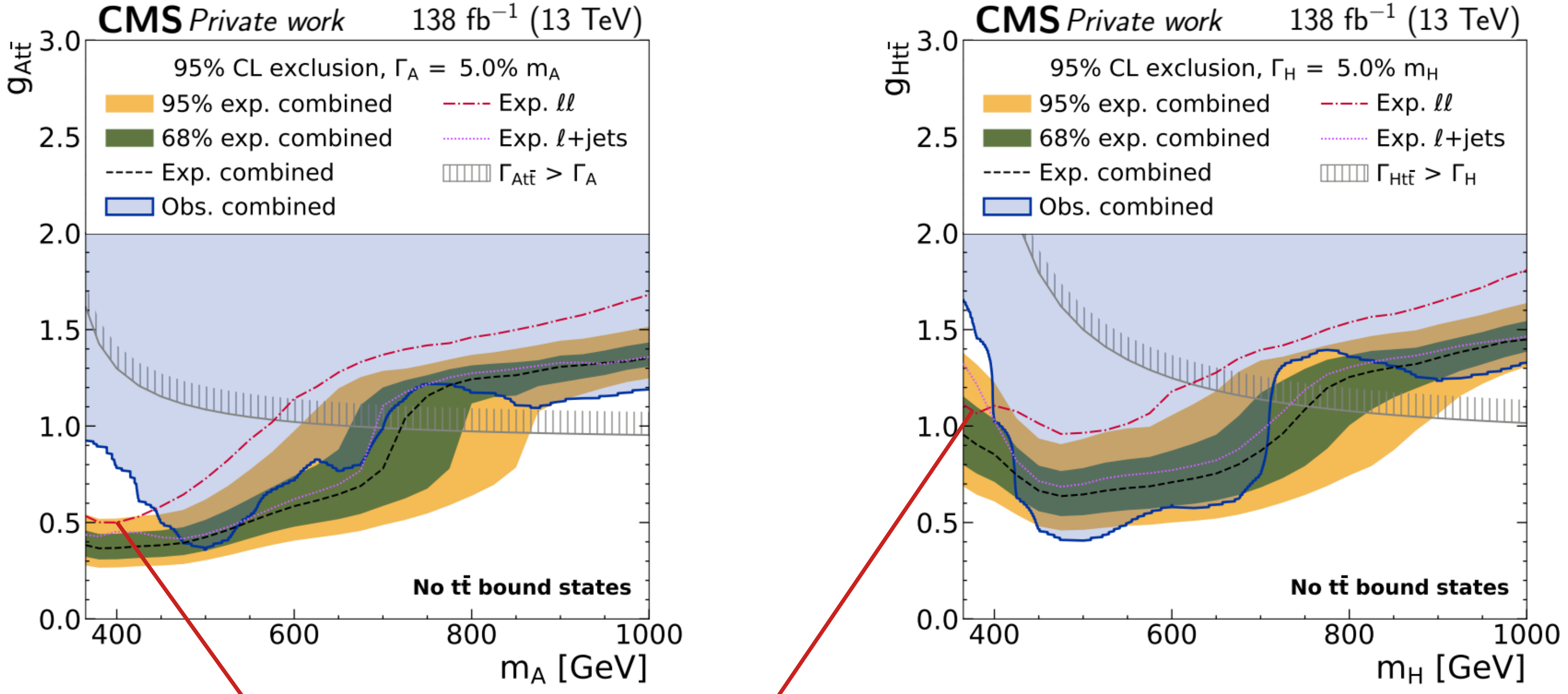
- Statistical combination with lepton+jets decay channel



Limits on A/H: combination

arXiv:2507.05119
submitted to *RoPP*

- Statistical combination with lepton+jets decay channel

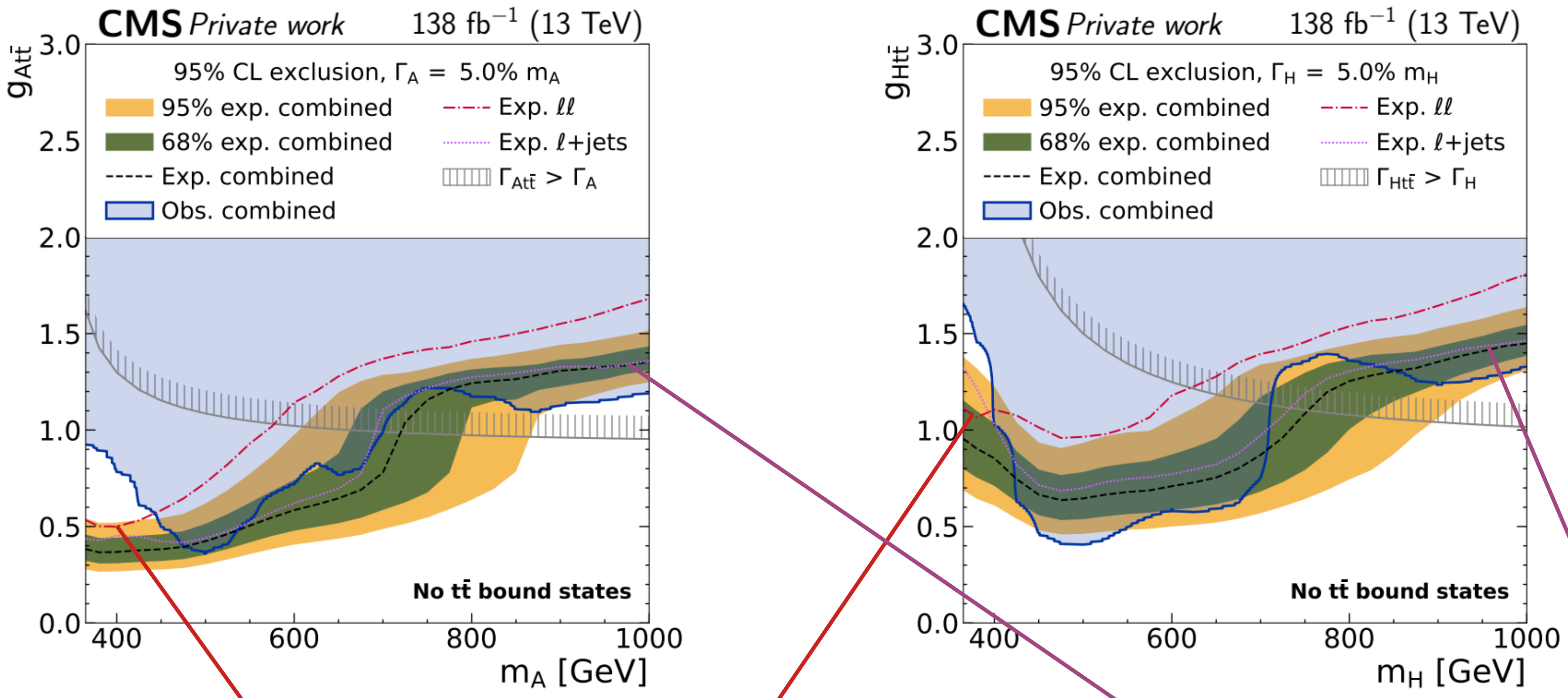


at low masses: dilepton channel comparable
or better than l+jets

Limits on A/H: combination

arXiv:2507.05119
submitted to *RoPP*

- Combine A/H limits with lepton+jets decay channel



at low masses: dilepton channel comparable or better than l+jets

at high masses: l+jets channel dominates

Axion-Like Particles

JHEP 12 (2024) 197

- Further interpretation: heavy Axion-Like Particles (ALPs)
  - could solve strong CP problem
  - if $m_{ALP} > 2\, m_t$: decay to $t\bar{t}$ just like pseudoscalar A
  - possible additional ALP-gluon coupling in Lagrangian

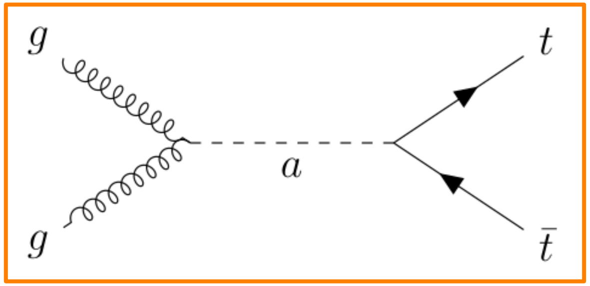
$$\mathcal{L}^{\mathrm{ALP}} = \frac{1}{2}(\partial_\mu a)(\partial^\mu a) + \frac{m_a^2}{2}a^2 - \boxed{c_{\tilde{G}}\frac{a}{f_a}G^a_{\mu\nu}\tilde{G}^{a\mu\nu}} + i c_t m_t \frac{a}{f_a}\bar{t}\gamma^5 t$$

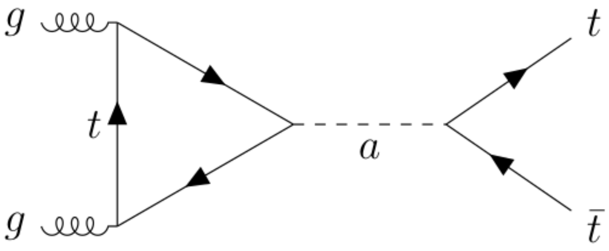
Axion-Like Particles

JHEP 12 (2024) 197

- Further interpretation: heavy Axion-Like Particles (ALPs)
  - could solve strong CP problem
  - if $m_{ALP} > 2\, m_t$: decay to $t\bar{t}$ just like pseudoscalar A
  - possible additional ALP-gluon coupling in Lagrangian

$$\mathcal{L}^{\mathrm{ALP}} = \frac{1}{2}(\partial_\mu a)(\partial^\mu a) + \frac{m_a^2}{2}a^2 - \boxed{c_{\tilde{G}}\frac{a}{f_a}G^a_{\mu\nu}\tilde{G}^{a\mu\nu}} + ic_t m_t \frac{a}{f_a}\bar{t}\gamma^5 t$$



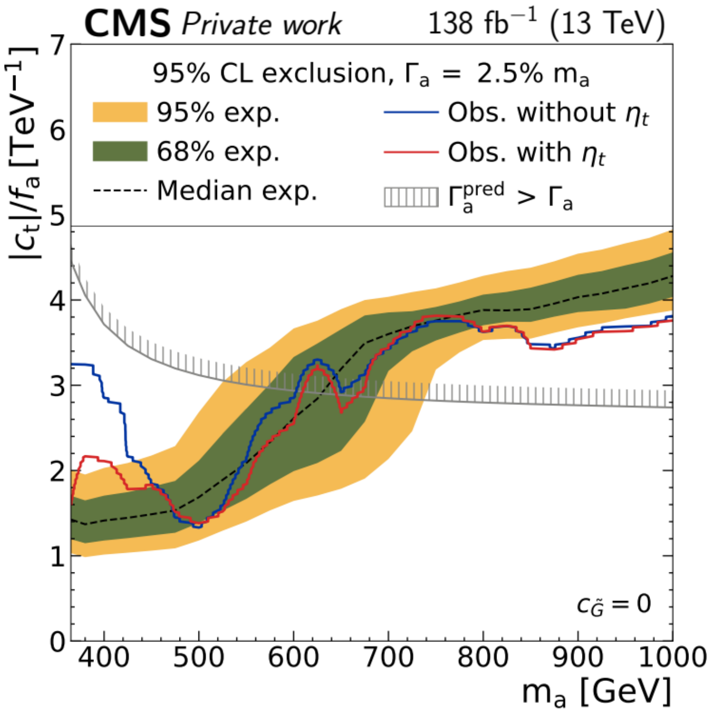


additional production diagram
→ different m_{tt} spectrum

Axion-Like Particles

JHEP 12 (2024) 197

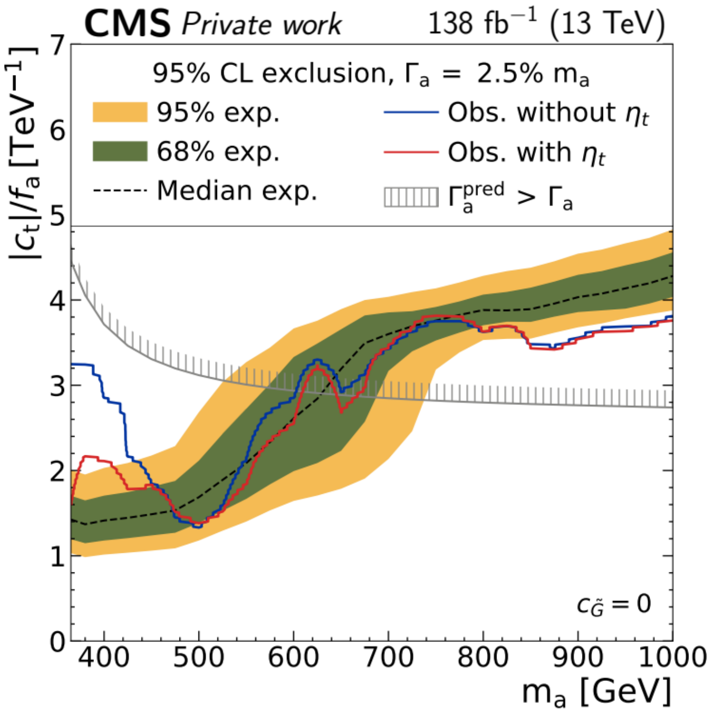
no ALP-gluon coupling $c_{\tilde{G}}$:
set experimental limits



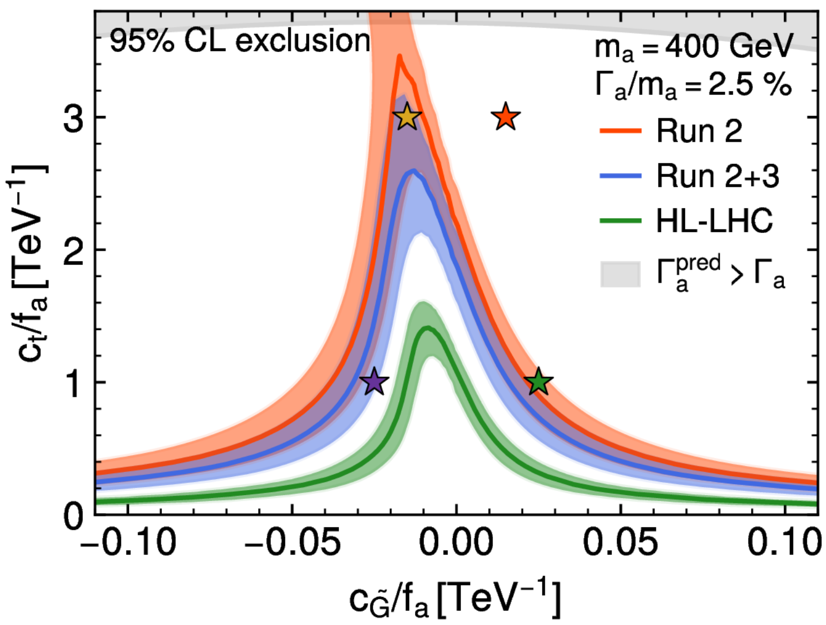
Axion-Like Particles

JHEP 12 (2024) 197

no ALP-gluon coupling $c_{\tilde{G}}$:
set experimental limits



with ALP-gluon coupling $c_{\tilde{G}}$:
projected limits for different
eras of LHC



Summary

- Early measurement of the inclusive $t\bar{t}$ cross section in Run 3
  - First public physics result of Run 3, only ~2 months after start of datataking
  - Required estimation of many needed corrections
  - Comparable precision to Run 2 measurements

JHEP 08 (2023) 204

- Search for new spin-0 states in $t\bar{t} \rightarrow$ dilepton events
  - Low-m_{tt} excess observed & interpreted as $t\bar{t}$ bound state η_t (> 5σ!)
  - Further interpretations as generic (pseudo)scalars in combination with lepton+jets channel or Axion-Like Particles

RoPP 88 087801 (2025)

arXiv:2507.05119 (submitted to *RoPP*)

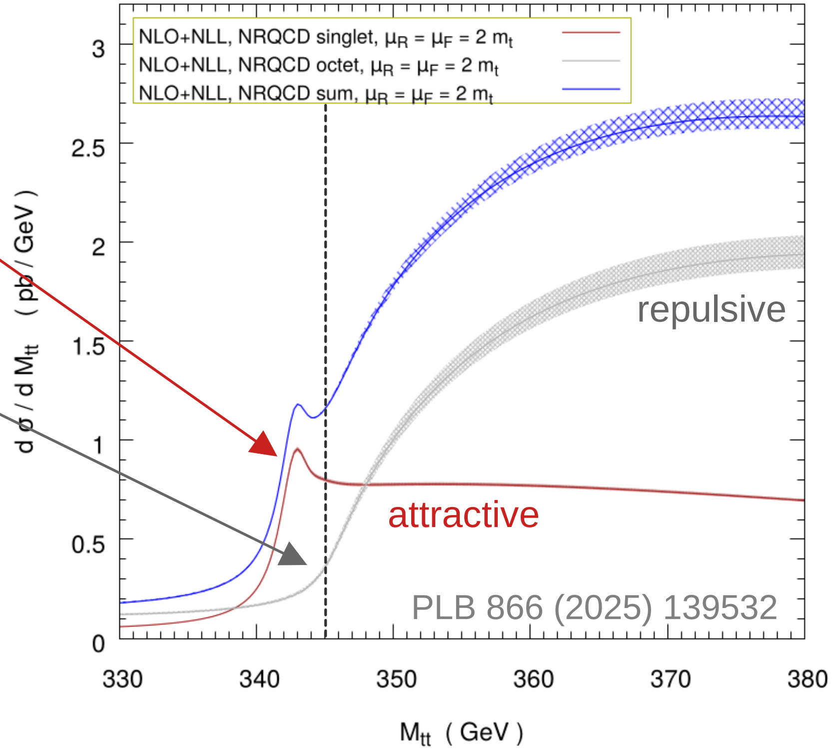
JHEP 12 (2024) 197

Backup

$t\bar{t}$ bound states: NRQCD

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

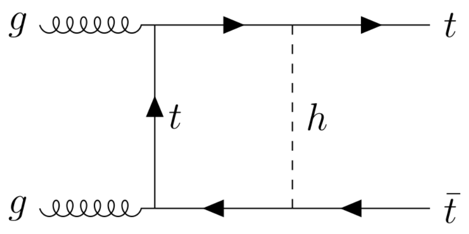
- Results of NRQCD calculations:
  - Color-singlet ($^1S_0^{[1]}$) - attractive
    $\rightarrow$ Peak below the $t\bar{t}$ threshold
    CP-odd / pseudoscalar spin state!
  - Color-octet ($^1S_0^{[8]}$ or $^3S_1^{[8]}$) - repulsive
    $\rightarrow$ Suppressed below the $t\bar{t}$ threshold
- Difficulties: matching to relativistic calculation
  soft gluon emissions, ...
- Exact lineshape and width below experimental resolution

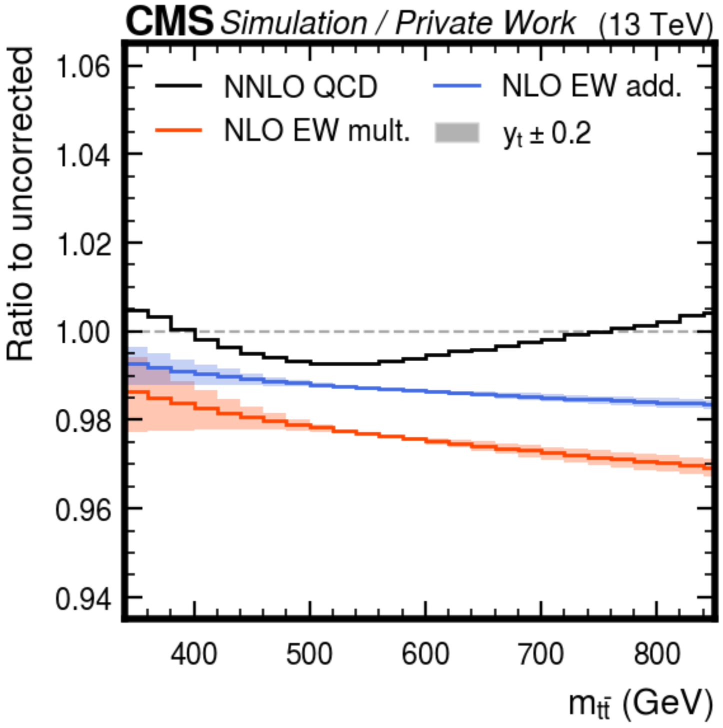


Background modeling

- 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 top scattering angle

    (EPJC 78 (2018) 537, EPJC 51 (2007) 37)





  - Normalize to NNLO+NNLL cross section

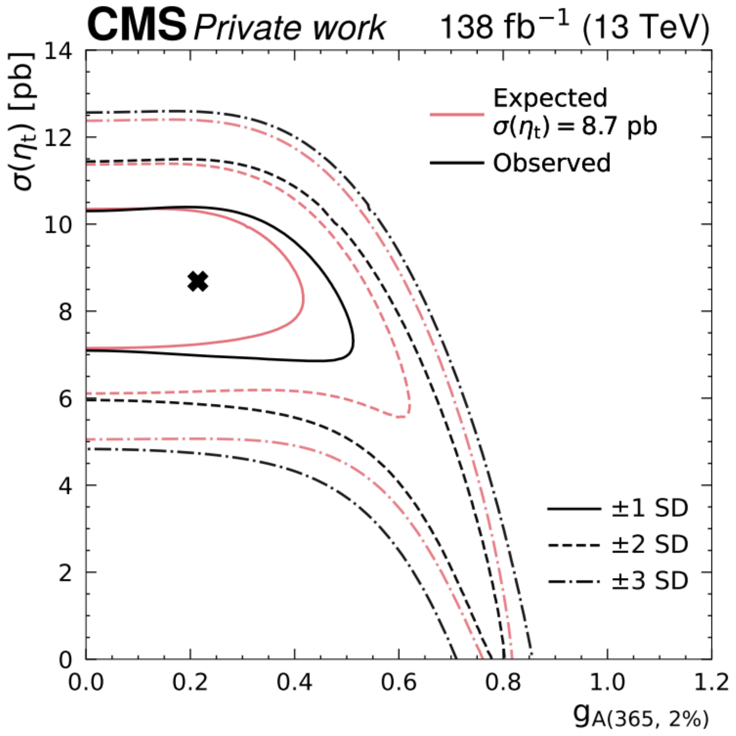
    (CPC 185 (2014) 2930)

$t\bar{t}$ bound state or BSM?

- same excess can be interpreted as (BSM) pseudoscalar A: best fit

$$m_{\mathrm{A}} = 365\,\mathrm{GeV}, \quad \Gamma_{\mathrm{A}}/m_{\mathrm{A}} = 2\%, \quad g_{\mathrm{At\bar{t}}} = 0.79$$

lowest mass point probed in the analysis!



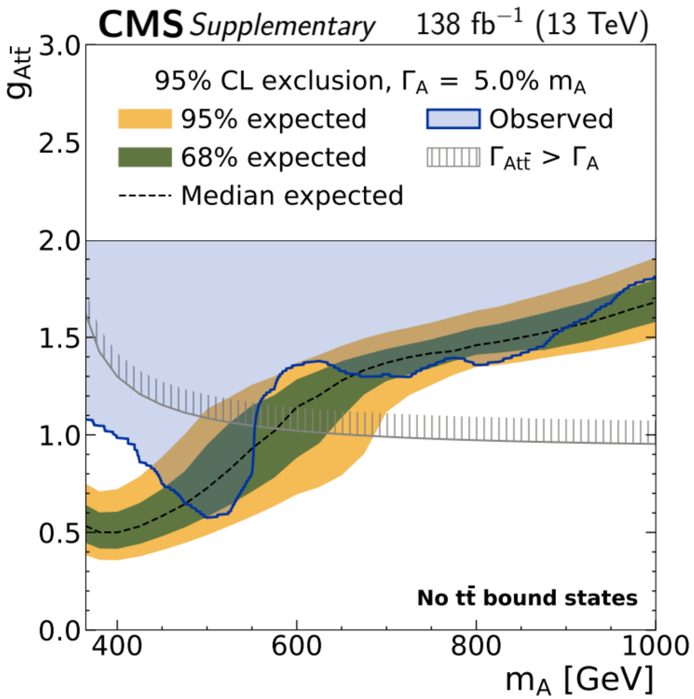
- in general: could be any combination from bound state effects and BSM
- fit prefers η_t bound state, but only by 1σ

Limits on A/H

- Set limits on couplings of A/H to top over large mass range
- Two scenarios on how to treat η_t:

no η_t:

excess visible in limits

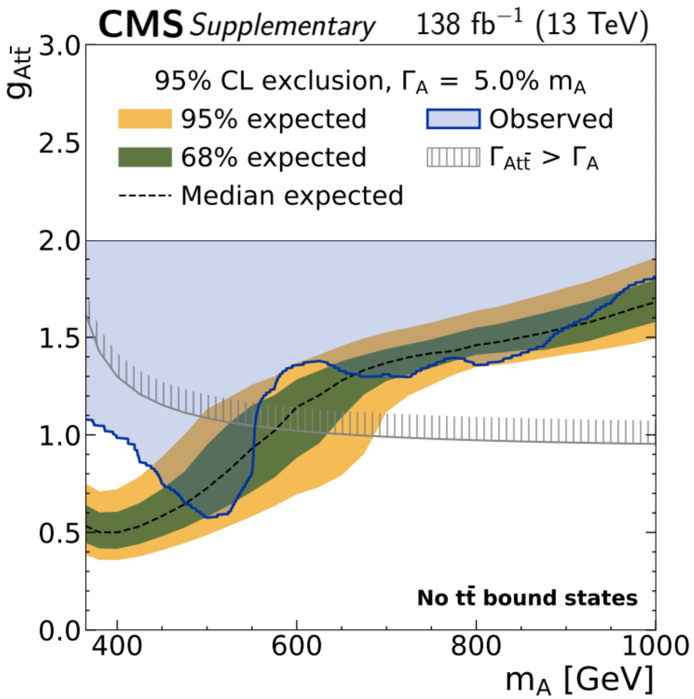


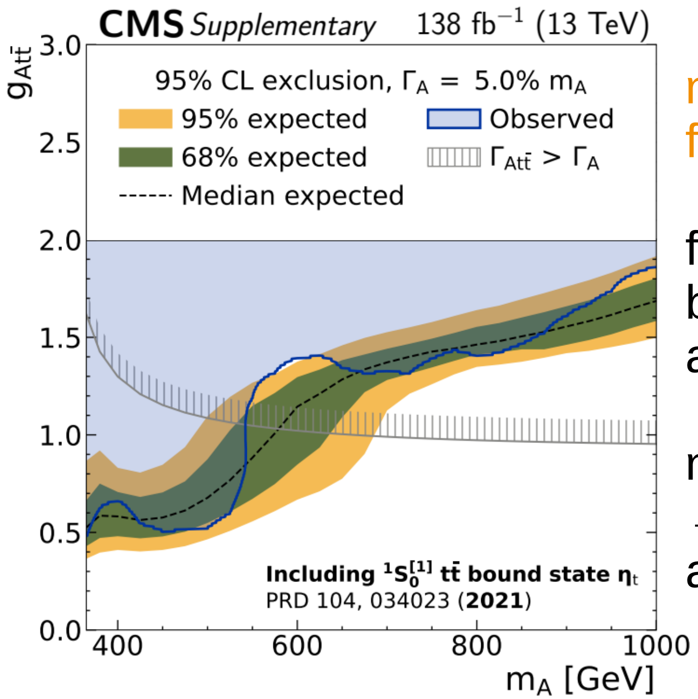
Limits on A/H

- Set limits on couplings of A/H to top over large mass range
- Two scenarios on how to treat η_t:

no η_t:

excess visible in limits





η_t as additional free-floating BG:

fit can choose between A/H and η_t

η_t is preferred → excess is absorbed

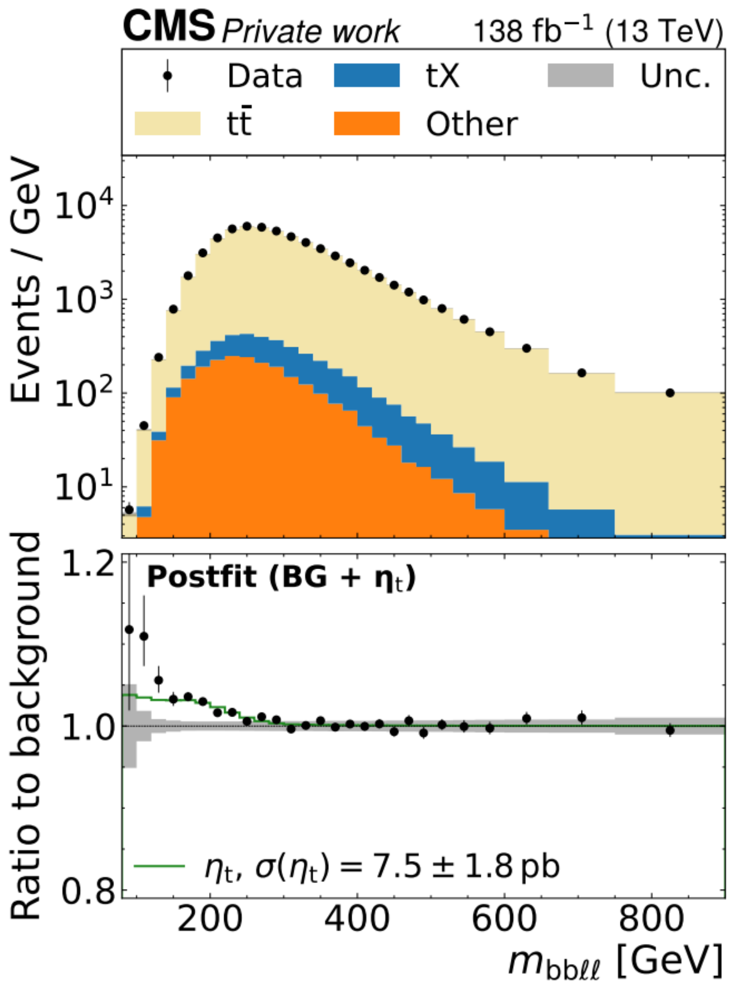
Checks of the result

How do we make sure the excess is real?

Check alternative variables: $m_{bb\ell\ell}$ instead of $m_{t\bar{t}}$

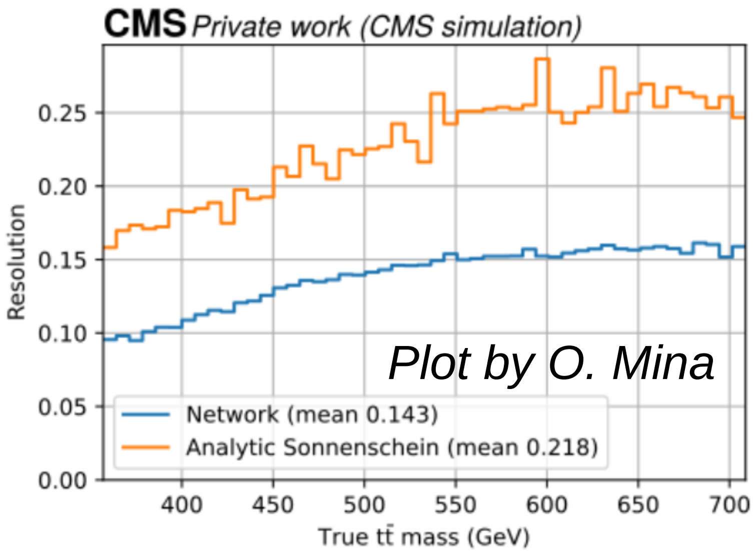
- Sidestep possible biases from $t\bar{t}$ reconstruction
- Repeat full fit: consistent with main result

… and many, many more



Future: ML for $t\bar{t}$ reconstruction

- Analytic dilepton $t\bar{t}$ reconstruction has several problems
  - Assumes top mass constraints – questionable in off-shell regions
  - b jet assignment done in simplistic way, etc...
- We should be able to do better!
- Use modern machine learning techniques: Lorentz-invariant transformers (L-GATr) [arXiv:2405.14806]
- Idea: also estimate confidence in the result? Could reject poorly-reconstructed events...



- Supervision of two summer student projects working on ML for tt reco.
  + one combined summer student+B.Sc. project working on the classic algorithm

How it was done: columnar analysis

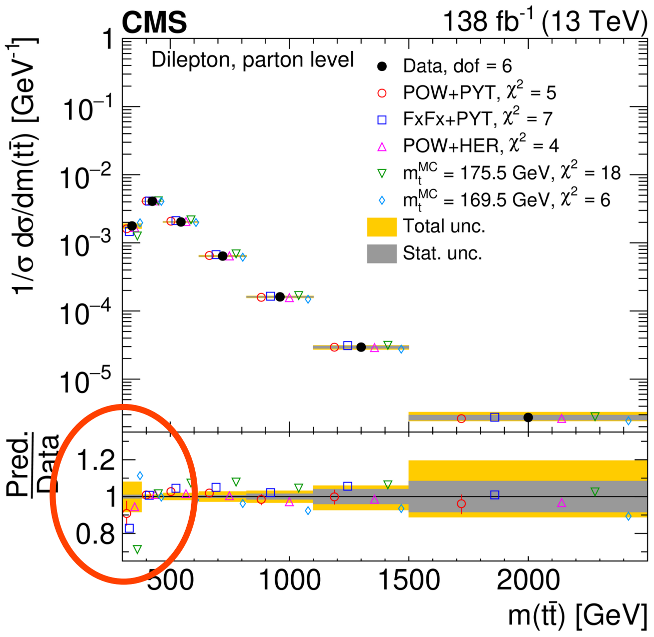
- Measurement was performed with pepper:
  fully python coffea-based columnar analysis framework
- Fully generic, starting from central nanoAOD
- Developed at DESY, by now also used at other CMS institutes

- I am one of two maintainers – implemented many new features & improvements
  - e.g. changes for Run 3, improved HTCondor submission, …
- Supervised combined summer student & Bachelor project:
  profile & improve performance of histogramming → up to 150% speedup

SM: $t\bar{t}$ differential measurements

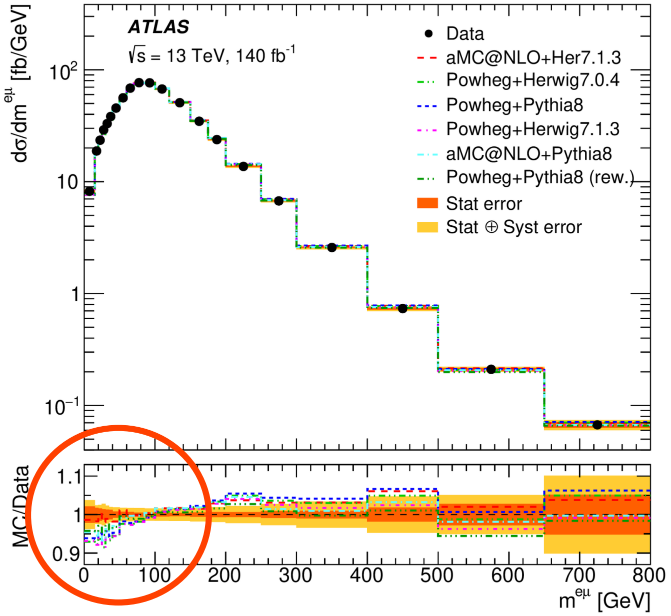
CMS ℓℓ - m_{tt}

CMS-PAS-TOP-20-006



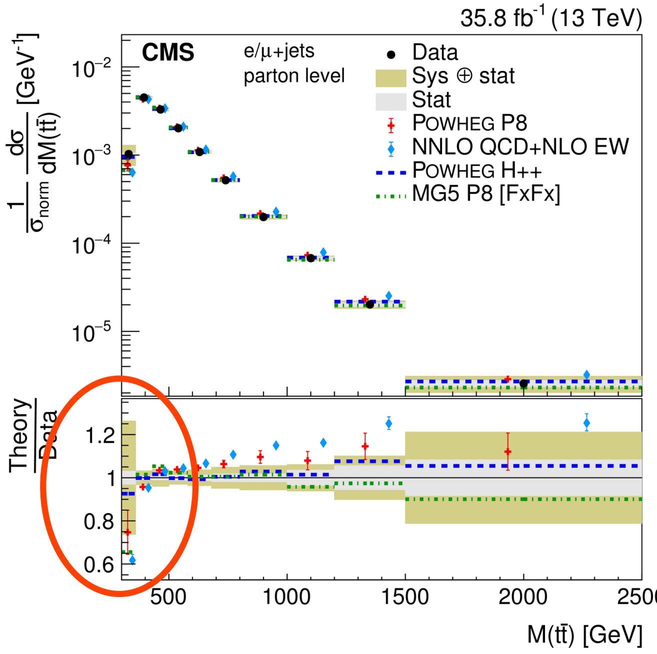
ATLAS eμ - $m_{e\mu}$

JHEP 07 (2023) 141



CMS ℓ+jets - m_{tt}

PRD 97 (2018) 112003



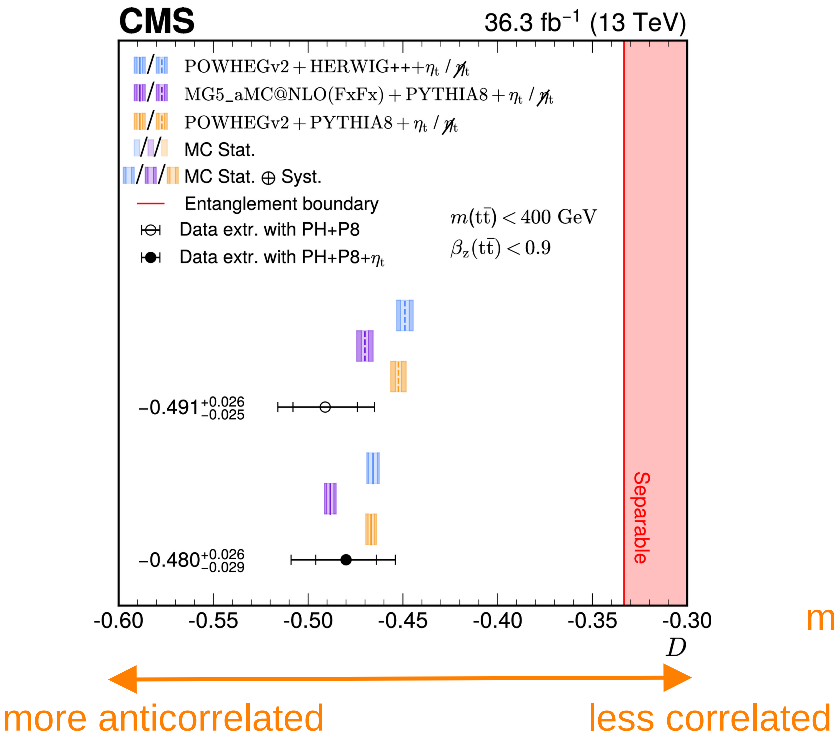
→ good description by theory except for excess in data in threshold region

SM: $t\bar{t}$ spin entanglement

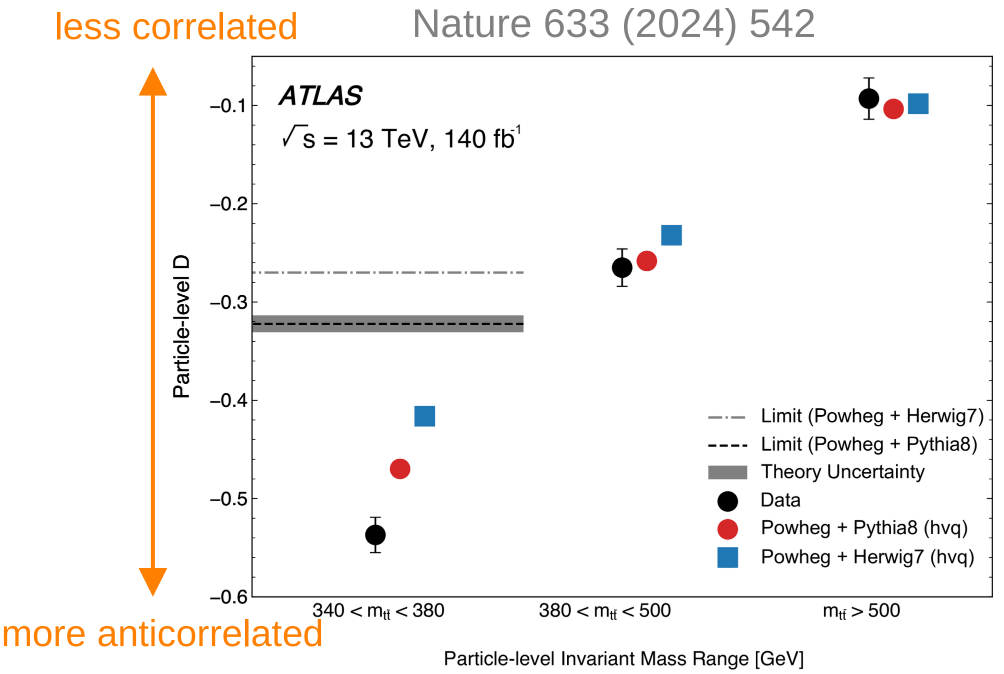
- Measured quantity: D "≈ strength of $t\bar{t}$ spin correlation"

somewhat oversimplified, more later…

ROPP 87 117801 (2024)

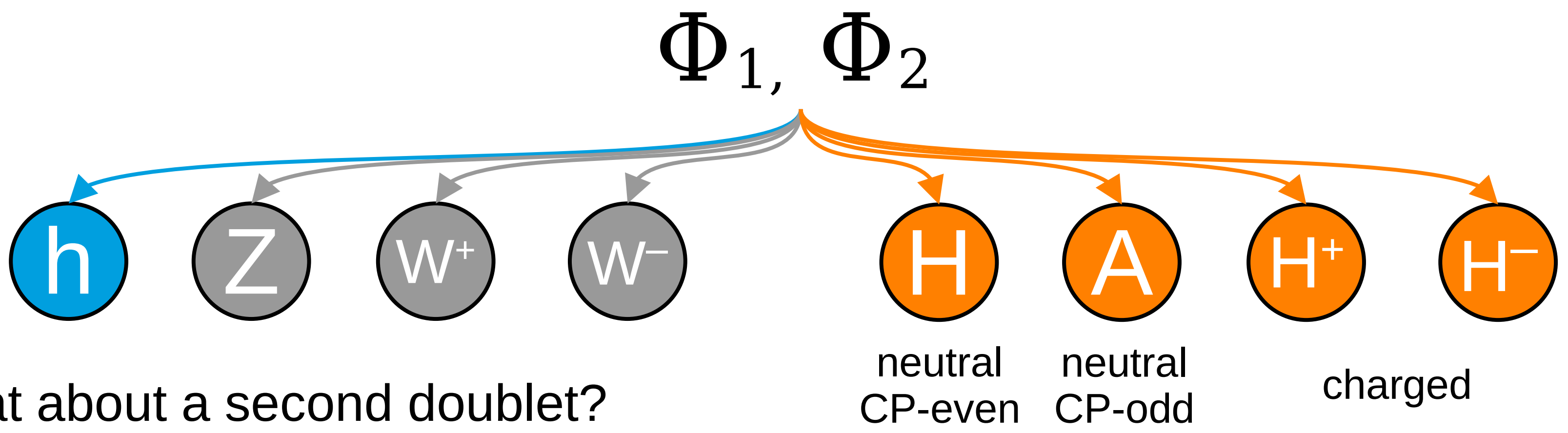


Nature 633 (2024) 542



BSM: Two-Higgs Doublet Model

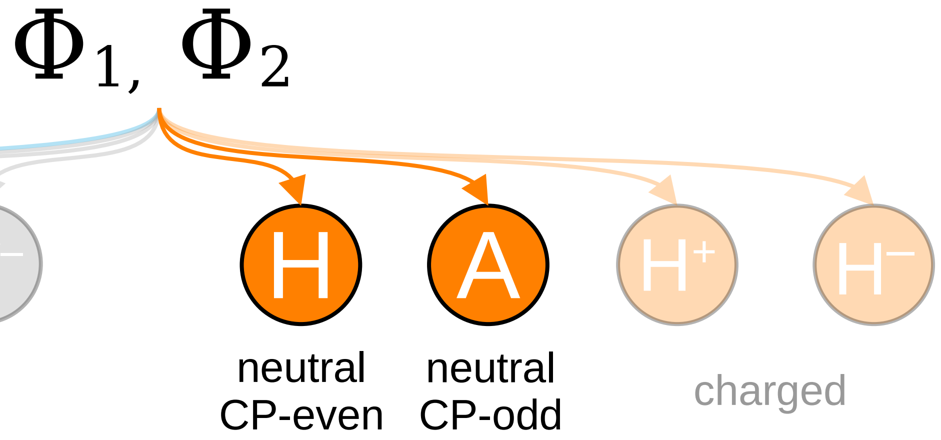
- SM Higgs: complex SU(2) doublet ϕ
  $\rightarrow$ becomes real scalar after symmetry breaking



- What about a second doublet?
  $\rightarrow$ Two-Higgs Doublet Model (2HDM)
  - Simplest UV-complete extension of the Higgs sector
  - 4 additional degrees of freedom: H, A, H^+, H^-
  - Only a starting point – e.g. included in SUSY

BSM: Two-Higgs Doublet Model

- SM Higgs: complex SU(2) doublet ϕ
  $\rightarrow$ becomes real scalar after symmetry breaking



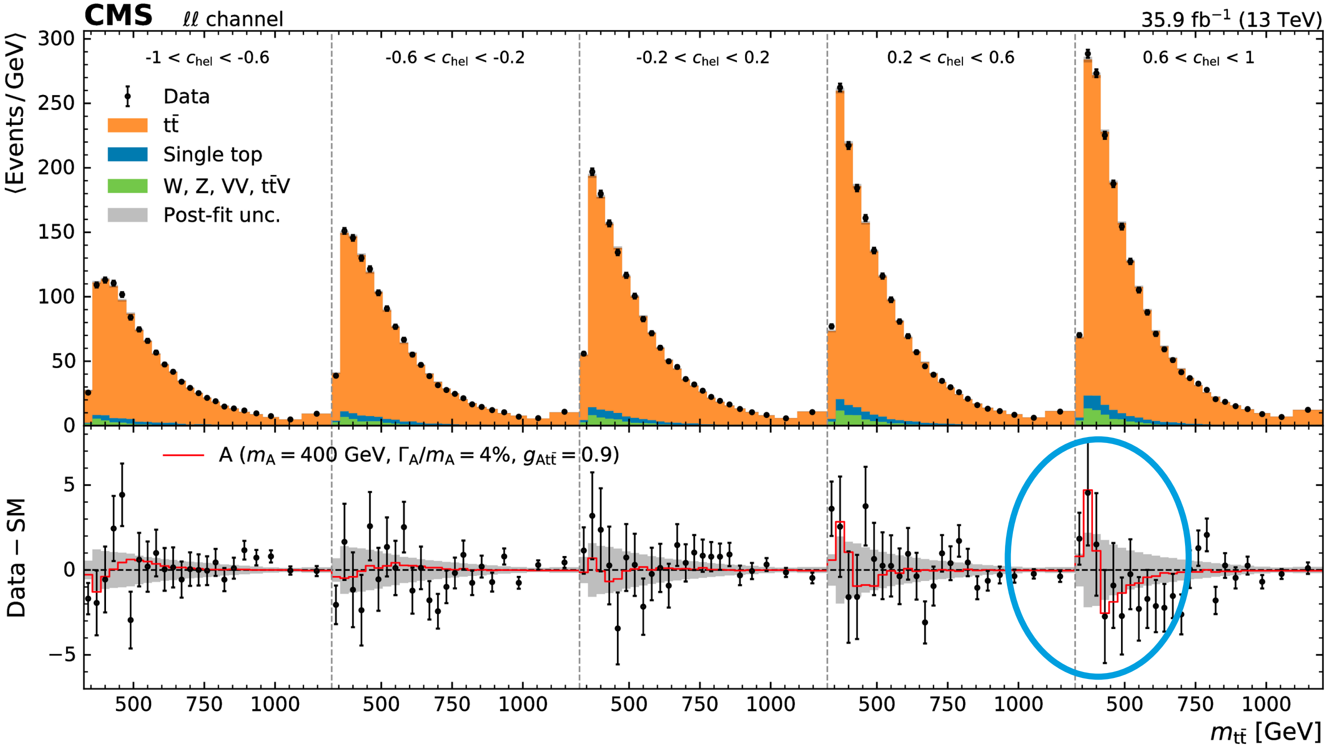
- Most models: couplings are Yukawa-like $\rightarrow$ largest coupling to top quark!
- For neutral states A and H: decay to $t\bar{t}$ dominates if $m_{A/H} > 2m_t$

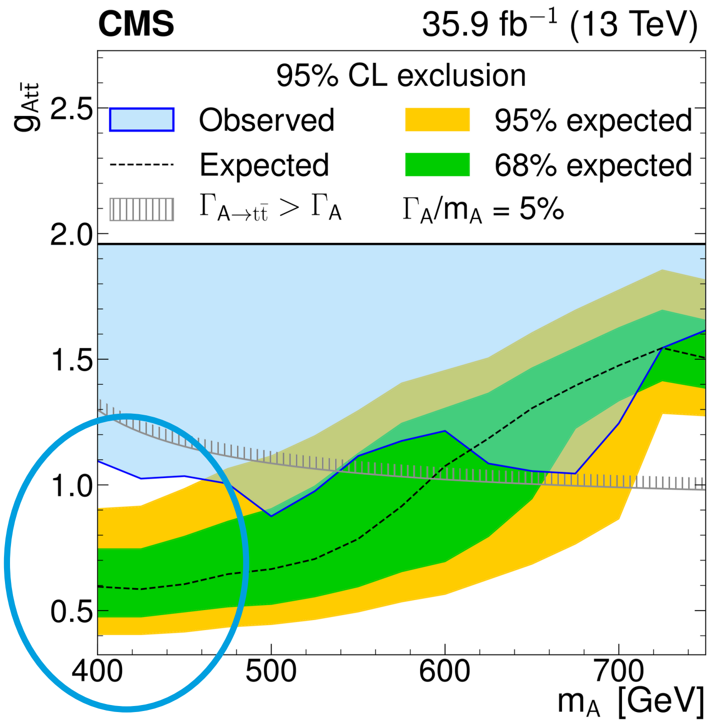
  $\rightarrow$ search in $t\bar{t}$ final states!

BSM: Previous work

Search for (pseudo)scalars in $t\bar{t}$ events with 2016 data

JHEP 04 (2020) 171





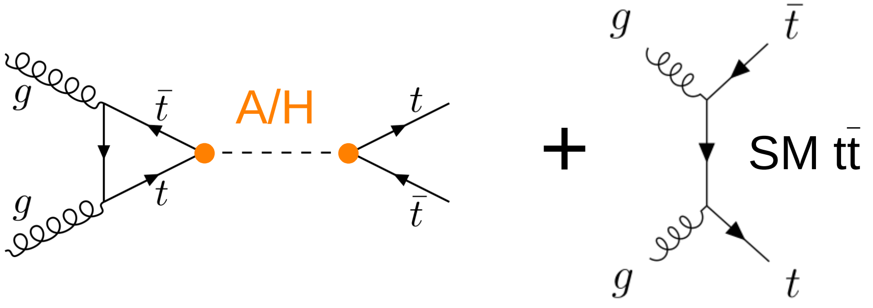
Excess for low pseudoscalar masses (~3σ local)

Signal modeling: A/H

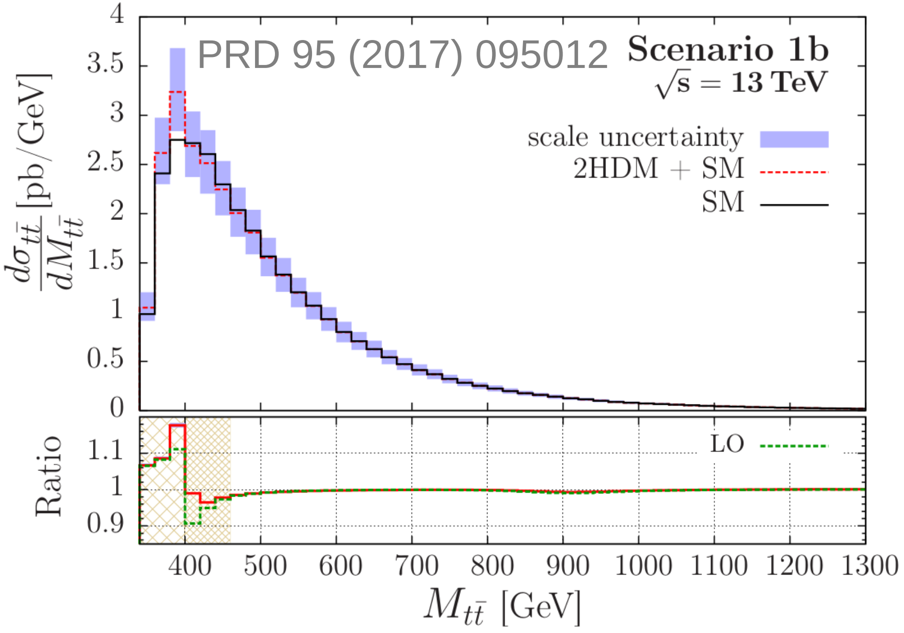
$$\mathcal{L}_A^{\text{int}} = i g_{\text{At}\bar{\text{t}}} \frac{m_\text{t}}{v} \bar{\text{t}} \gamma_5 \text{t} \text{A}$$

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

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



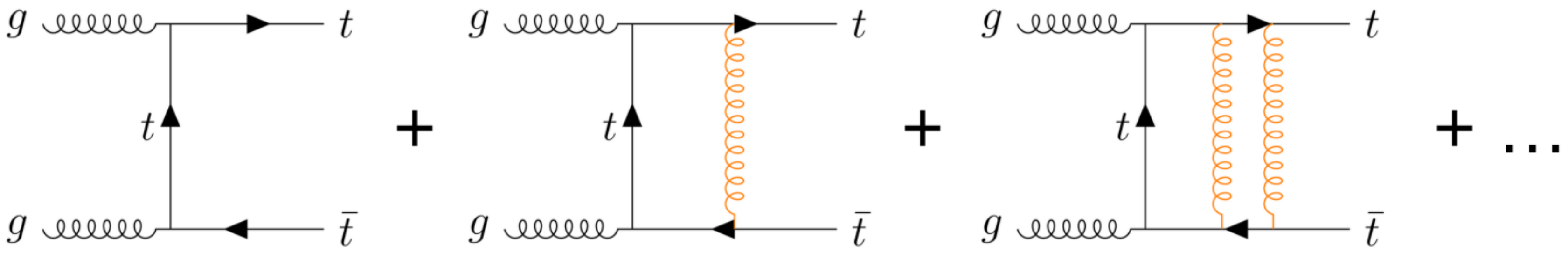
- Same final state as SM $t\bar{t}$ → interference → peak-dip structure in m_{tt}
- Free parameters: masses, widths, coupling modifiers g_A / g_H



Modeling: $t\bar{t}$ bound states

- How to calculate bound states of heavy quarks?
  $\rightarrow$ resummation of soft gluons (Coulomb singularities) in α_S/β

top velocity



- equivalent to solving non-relativistic Schrödinger equation with QCD potential

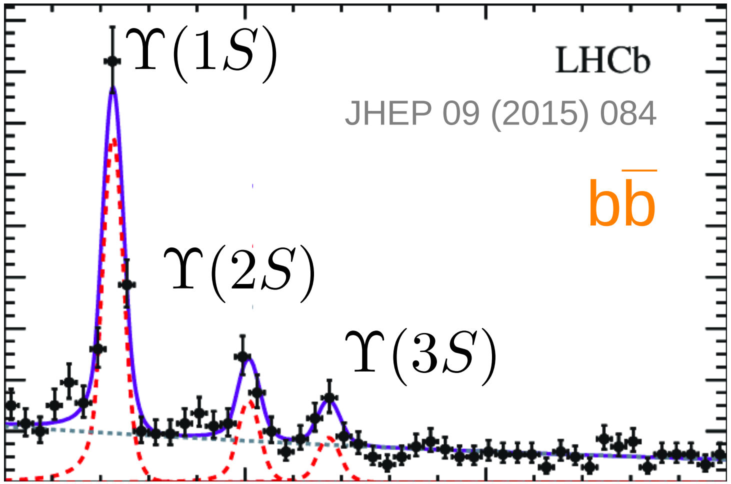
color factor

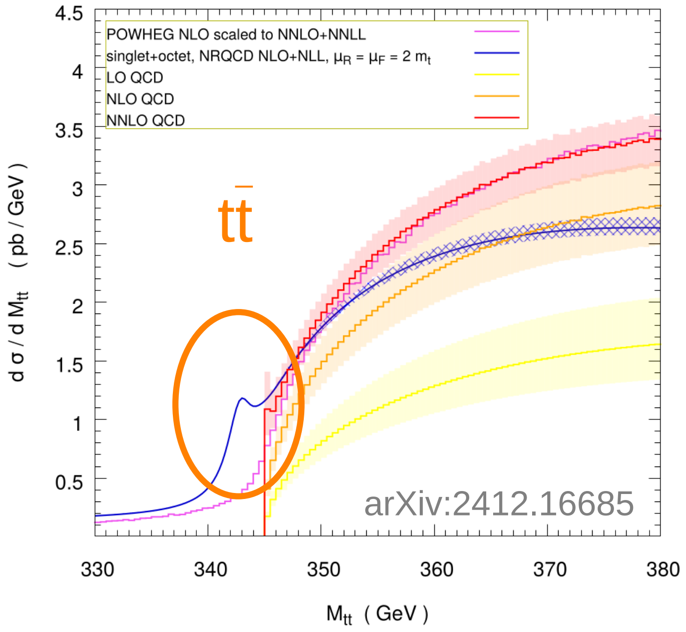
momentum space
like e.g. hydrogen atom!

$$V_{\text{QCD}}(|\vec{q}|) = -\frac{4\pi C^{[1,8]}\alpha_S(|\vec{q}|)}{\vec{q}^2} + \mathcal{O}(\alpha_S^2)$$

Modeling: $t\bar{t}$ bound states

- Need to consider finite top width
  $\rightarrow$ one broad peak instead of multiple narrow bound states





➔ "virtual bound state" or "quasi-bound state"

- Short top lifetime $\rightarrow$ bound state disassociates by one top decaying to Wb

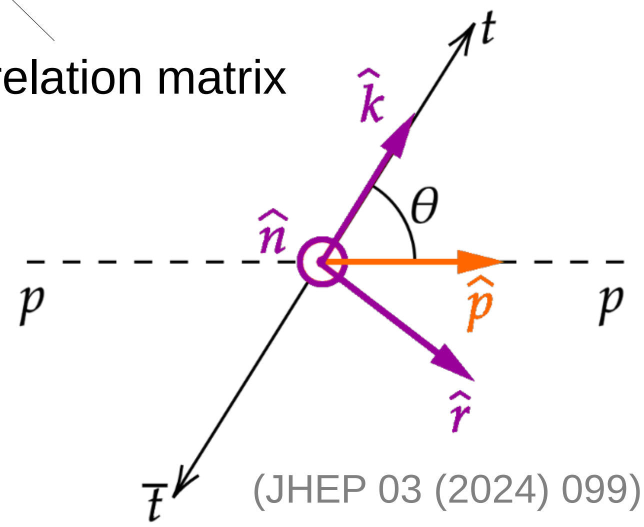
Spin density matrix

- 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)
- Encoded in spin density matrix:

$$\mathbf{R} = A + B_i\sigma_i + \overline{B}_i\overline{\sigma}_i + \sigma_i C_{ij}\overline{\sigma}_j$$

cross section — polarization vectors — correlation matrix

- Choose helicity basis $\{\hat{k}, \hat{r}, \hat{n}\}$:
  - $\hat{k}$: direction of flight of the top quark
  - $\hat{r}$ and $\hat{n}$: orthogonal to $\hat{k}$



(JHEP 03 (2024) 099)

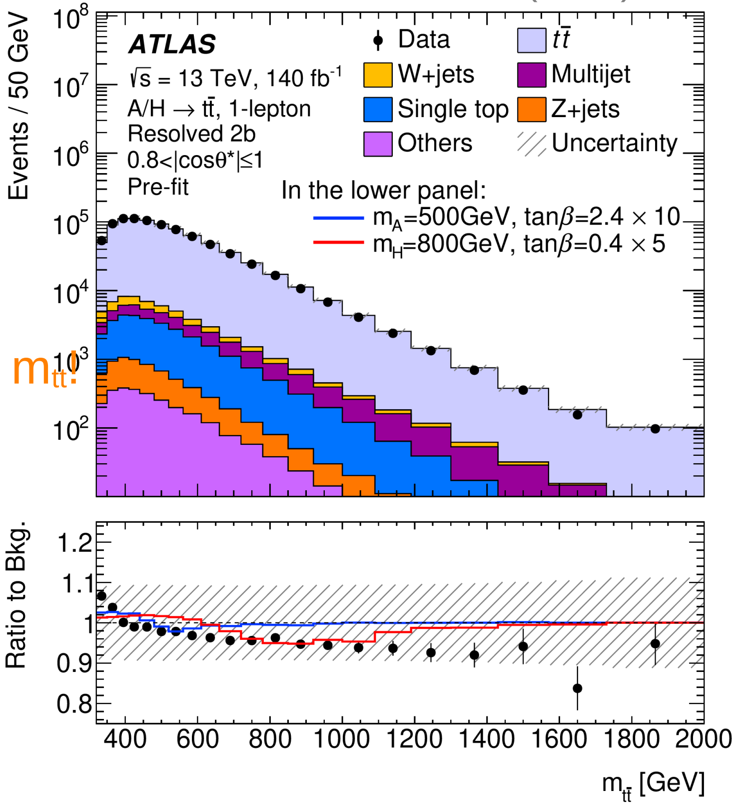
Comparison with ATLAS

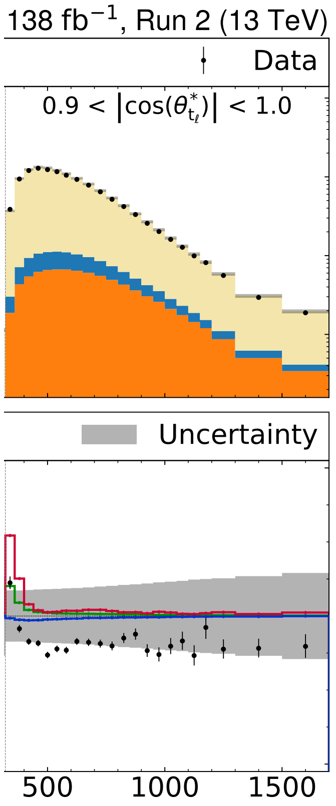
- Different channel definitions in ATLAS and CMS
- ℓ+jets resolved:
  - ATLAS: 1ℓ, 1b, ≥4 jets
  - CMS: 1ℓ, 2b, 3 jets
  - both: 1ℓ, 2b, ≥4 jets
    → compare pre-fit distributions!

Comparison with ATLAS

- Different channel definitions in ATLAS and CMS
- ℓ+jets resolved:
  - ATLAS: 1ℓ, 1b, ≥4 jets
  - CMS: 1ℓ, 2b, 3 jets
  - both: 1ℓ, 2b, ≥4 jets
    → compare pre-fit distributions!
  - Similar prefit excess in data at low $m_{t\bar{t}}$!

JHEP08 (2024) 013





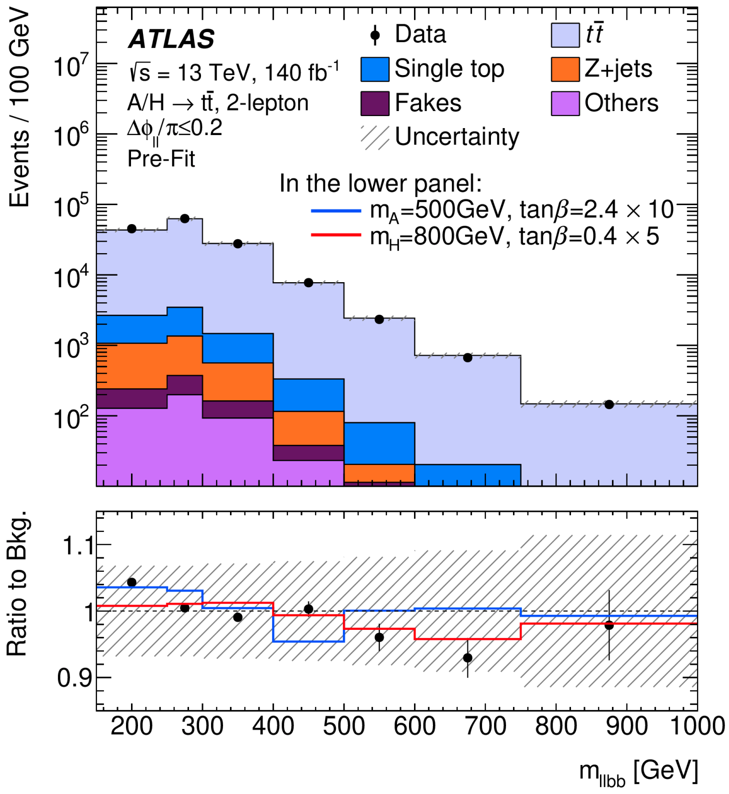
Comparison with ATLAS

- Different channel definitions in ATLAS and CMS

- ℓ+jets resolved:
  - ATLAS: 1ℓ, 1b, ≥4 jets
  - CMS: 1ℓ, 2b, 3 jets
  - both: 1ℓ, 2b, ≥4 jets
    → compare pre-fit distributions!
  - Similar prefit excess in data at low m_{tt}!

- dilepton: very different strategy
  - CMS: reconstruct m_{tt} x c_{hel} x c_{han} drives sensitivity at the $t\bar{t}$ threshold!
  - ATLAS: no top quark reconstruction instead: $m_{\ell\ell bb}$ x $\Delta\phi_{\ell\ell}$

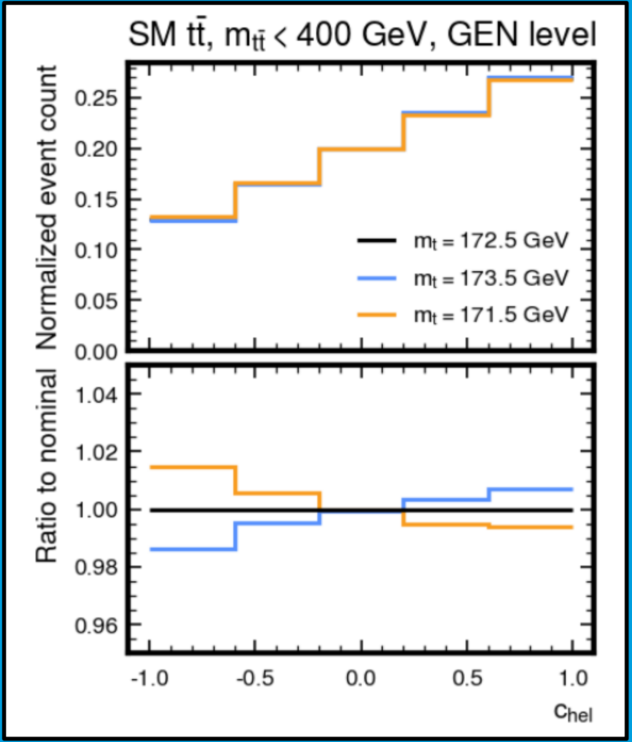
JHEP08 (2024) 013



Comparison with ATLAS

- From internal studies:
  inclusion of **spin correlations** in CMS helps to disentangle signal and systematic uncs.
- e.g. for **downwards shift of top mass**:
  - More events at threshold → like signal :(
  - BUT less spin correlation → unlike signal :)
- Similar for many other uncertainties
  e.g. Pythia vs. Herwig in the $t\bar{t}$ BG

➔ (one) reason why **CMS dilepton is more sensitive at the $t\bar{t}$ threshold**



Comparison with ATLAS

- Similar prefit excess in comparable lepton+jets channel
- Very different strategy in dilepton channels
  - drives the CMS sensitivity at the threshold
  - subdominant for ATLAS (no top reconstruction)
- Some different uncertainty treatment

- We are comparing in detail in the LHC Top Working Group!

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

$$\frac{(R^{\geq 1b}_{in/out})_{data}}{(R^{\geq 1b}_{in/out})_{MC}} = \frac{(R^{0b}_{in/out})_{data}}{(R^{0b}_{in/out})_{MC}}$$

→

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

with $N_{data} = N^{\ell\ell}_{data} - 0.5 N^{e\mu}_{data} k_{\ell\ell}$, 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

$$\sigma^{\text{rew.}} = \underbrace{\sigma^{\text{LO EW}}_{\text{NLO QCD}}}_{\text{Powheg}} \times \frac{\overbrace{\sigma^{\text{NLO EW}}_{\text{LO QCD}}}^{\text{Hathor}}}{\underbrace{\sigma^{\text{LO EW}}_{\text{LO QCD}}}_{\text{MadGraph}}}$$

- 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