Impact of QCD and PDF uncertainties on BSM searches (and Higgs)

Miha Muškinja *on behalf of ATLAS and CMS Collaborations*

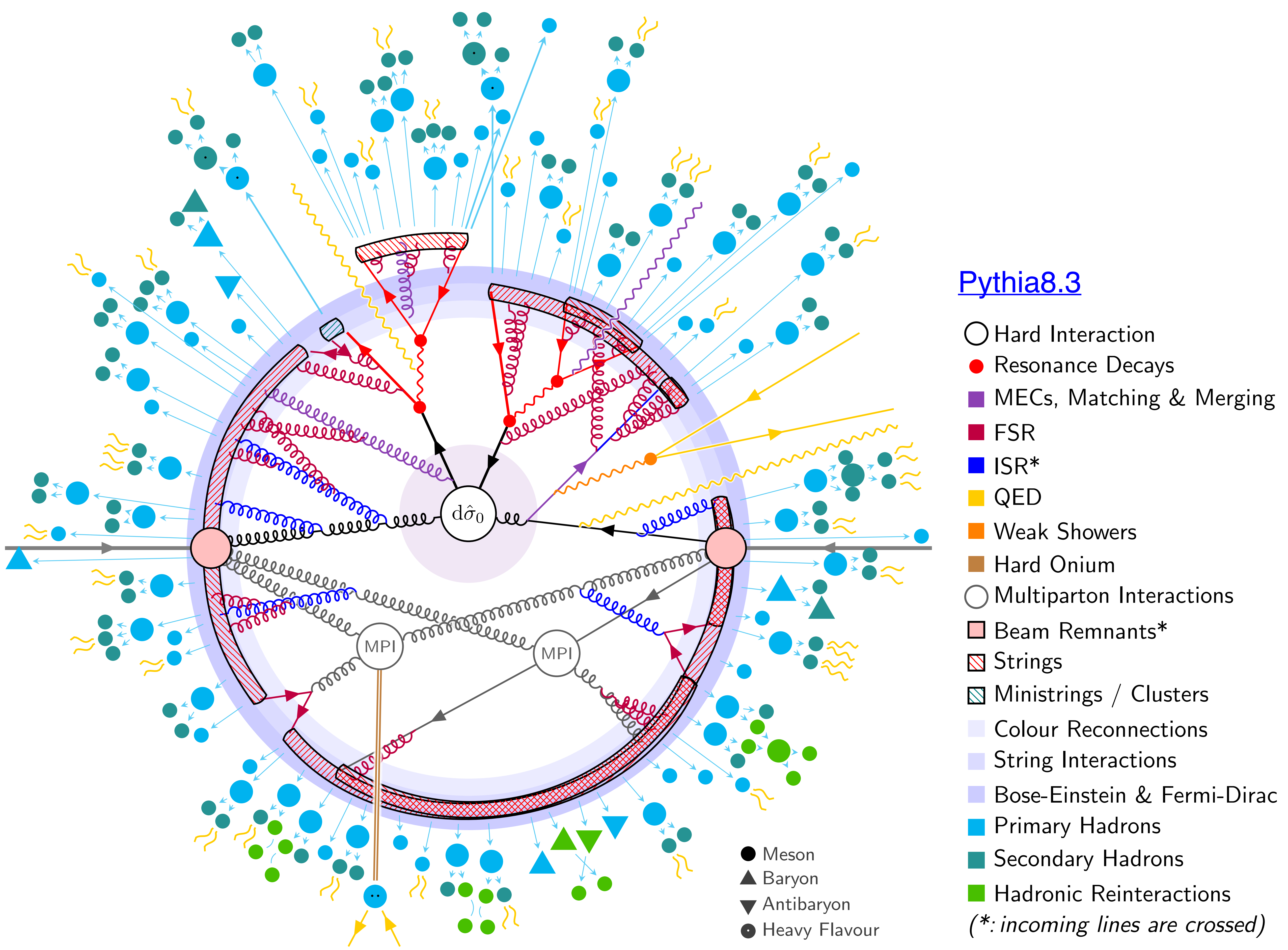
QCD@LHC 2024

Thursday, October 10, 2024

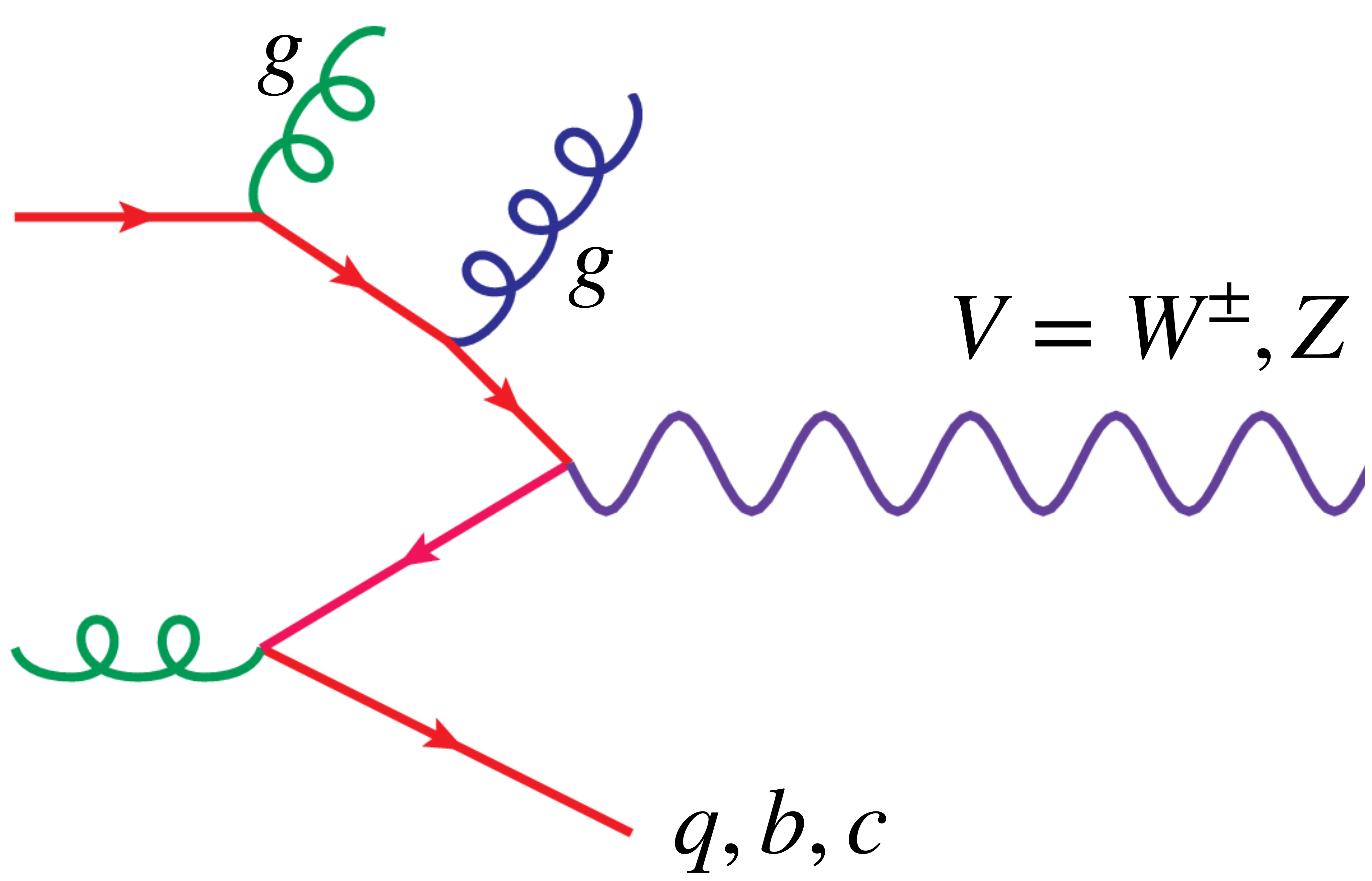
UNIVERSITY OF LJUBLJANA
FMF
Faculty of Mathematics and Physics
Jožef Stefan Institute, Ljubljana, Slovenia
ATLAS EXPERIMENT
"YOU ARE WELCOME HERE!"

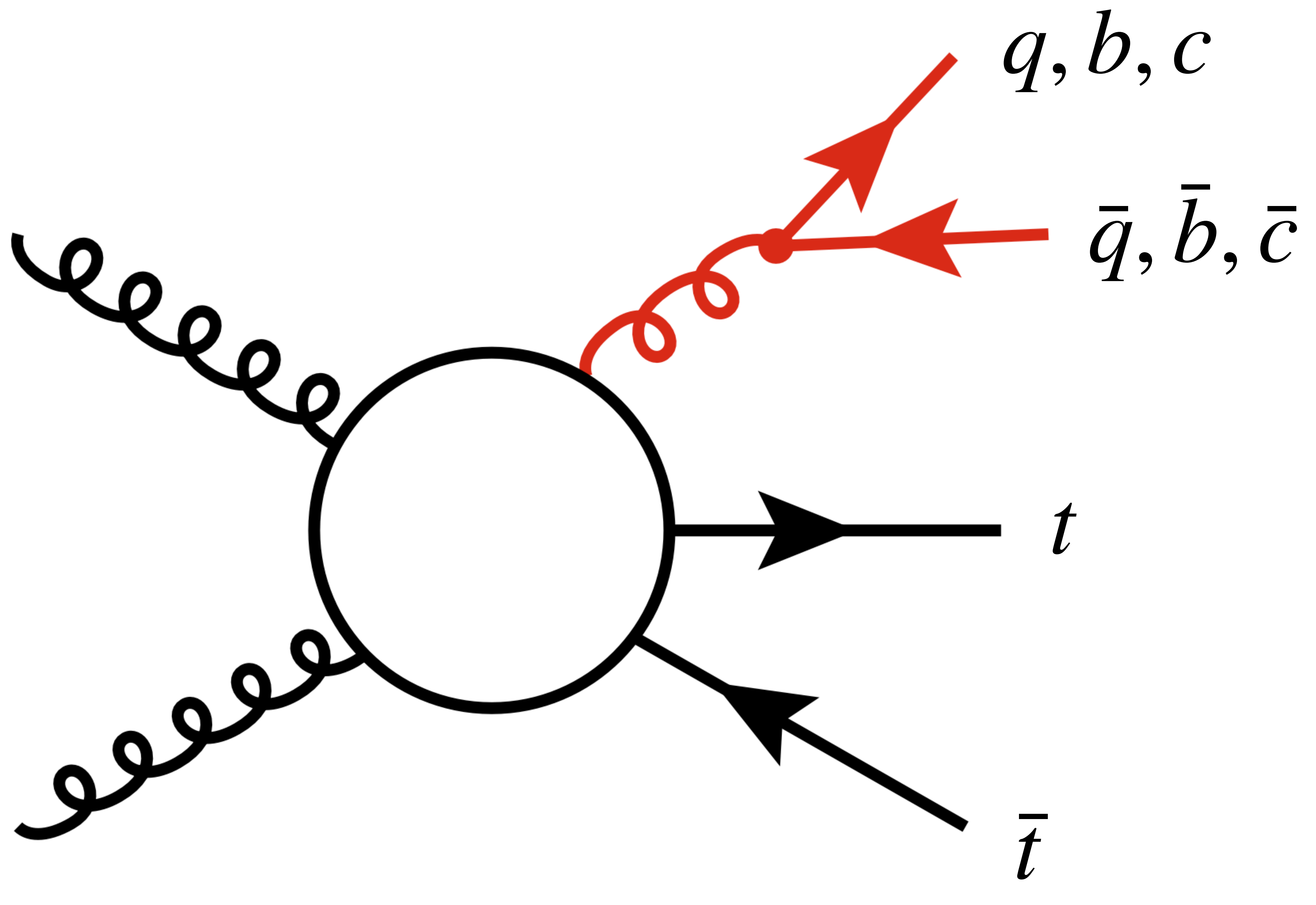
- Analytical QCD calculations by theorists are essential to predict signal and background processes
- **Monte Carlo event generators** are crucial for translating these theoretical calculations into realistic event simulations, accounting for the restricted phase spaces relevant to LHC experiments
  - Many good talks on state-of-the-art developments throughout the week!

- **The hard interaction**: initial partons selected with Parton Distribution Functions (PDFs); kinematics of outgoing partons based on Matrix Element calculations in perturbative QCD
- **Parton shower**: Initial State Radiation (ISR) of additional particles (partons, photons) using numerical resummation of soft and collinear gluon emission; Final State Radiation (FSR)
- **Matching**: Consistently combining fixed-order calculations with parton shower
- Hadronization, Hadron decays, etc…



- V+jets (V+HF) and $t\bar{t}$ (+jets) are the most common backgrounds in searches at the LHC
- For inclusive topologies ATLAS / CMS typically use NLO generators plus PS (e.g. **Powheg+Pythia8**)
- For final states with many jets or heavy-flavor jets we normally use multi-jet merged setups
  - E.g. **MadGraph+Pythia8** with **FxFx** merging at NLO or **Sherpa** with **MEPS@NLO** merging
- Fixed-order calculations for heavy-flavor final states now available at NNLO (e.g. Rene's talk)
  - Hints that NLO predictions might underestimate scale uncertainties— careful treatment needed
- Standard Model precision measurements play a crucial role so that we can tune and validate simulations





- Backgrounds normalized to data in Control Regions (**CR**) and extrapolated to **SR** using MC simulation*
- We need to account for the “extrapolation uncertainty” as the modeling may differ between CR and SR
- Uncertainties can be reduced as long as shapes of features used to define CR and SR are well modeled

$N_B^0(\mathrm{CR})$ … Nominal MC prediction for CR

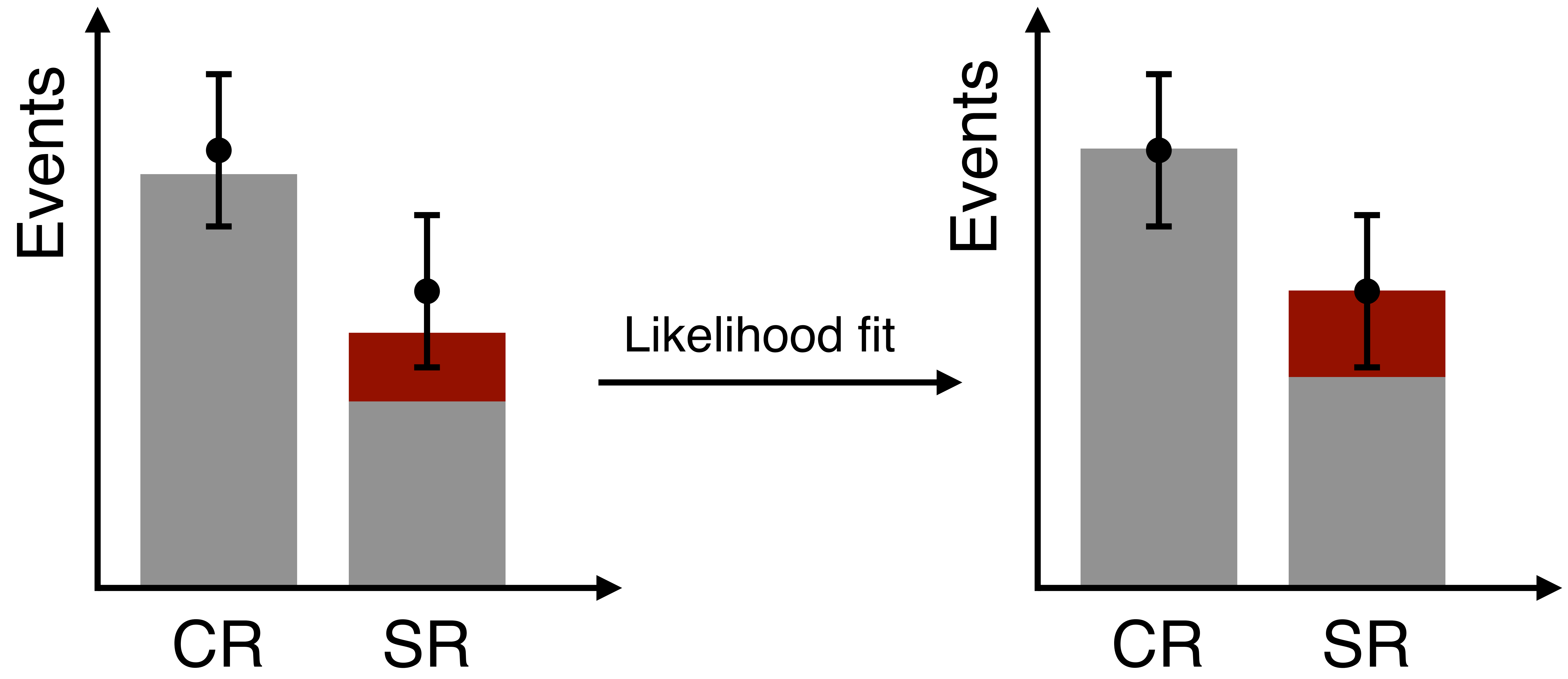
$N_B^0(\mathrm{SR})$ … Nominal MC prediction for SR

$N_B^i(\mathrm{CR})$ … MC theory variation “i” for CR

$N_B^i(\mathrm{SR})$ … MC theory variation “i” for SR

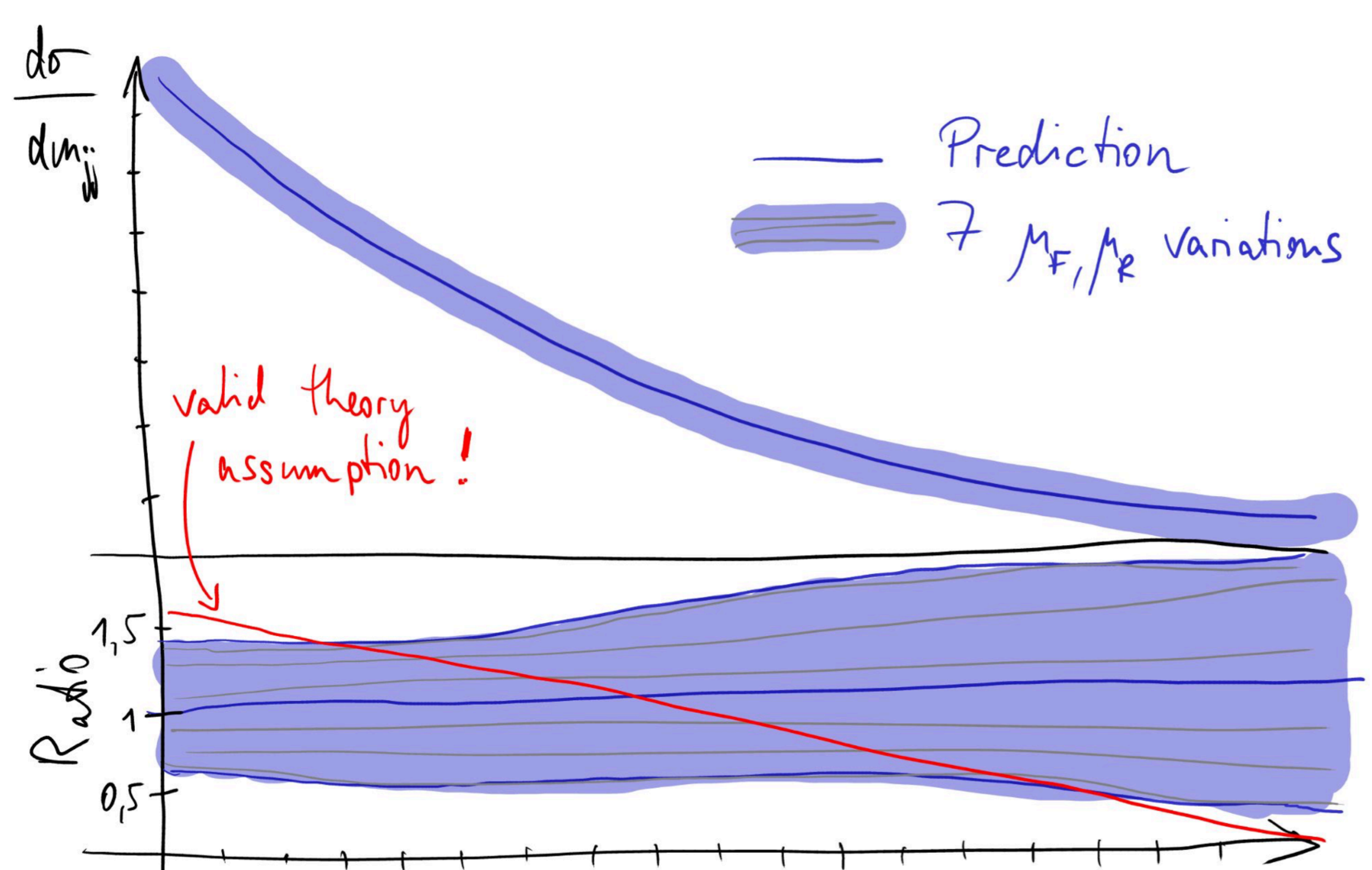
$$\mu^i = \frac{N_B^{\mathrm{data}}(\mathrm{CR})}{N_B^i(\mathrm{CR})}$$

$$\sigma^i_{\mathrm{rel.}} = \frac{\mu^i N_B^i(\mathrm{SR})}{\mu^0 N_B^0(\mathrm{SR})} = \frac{\dfrac{N_B^i(\mathrm{SR})}{N_B^i(\mathrm{CR})}}{\dfrac{N_B^0(\mathrm{SR})}{N_B^0(\mathrm{CR})}}$$

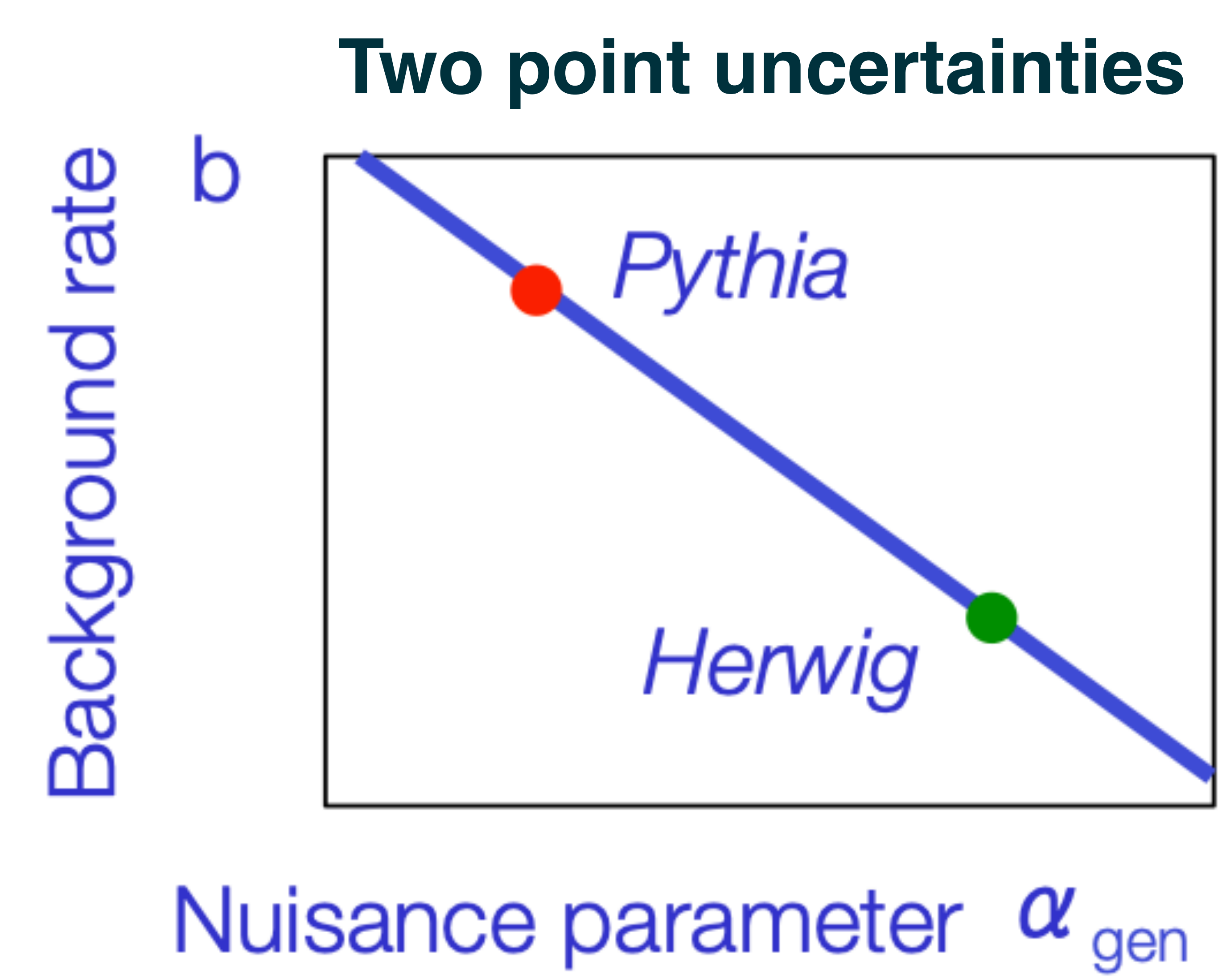


*: In likelihood fits backgrounds typically have normalization factors and are constrained both in CRs and SRs

- **Scale uncertainties** (μ_R and μ_F variations): we are not given a correlation scheme and LO / NLO generators may underestimate scale uncertainties
  - Taken to the extreme when we fit MVA outputs with 100% correlated scale uncertainties
  - Conservative approach is recommended, e.g. like in Lindert et al, 1705.04664 (see next slides)
- **Two point uncertainties**: taking the difference between two MC simulations and profiling it
  - No physical meaning to such variations (e.g. what is a -0.7σ variation of Pythia vs Herwig)
  - Not ideal because it usually accounts for too many changes at once (e.g. both ME + PS)
  - Differences in non-perturbative effects may arise from different tuning strategies rather actual modeling



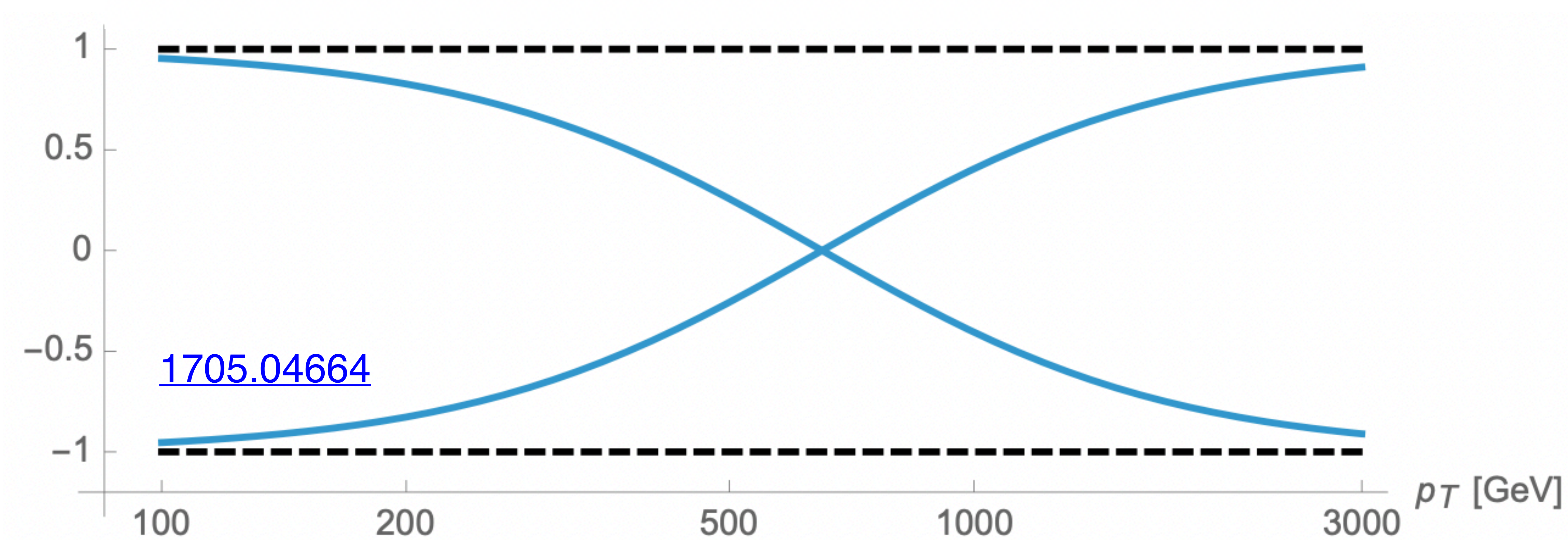
[Frank Siegert]

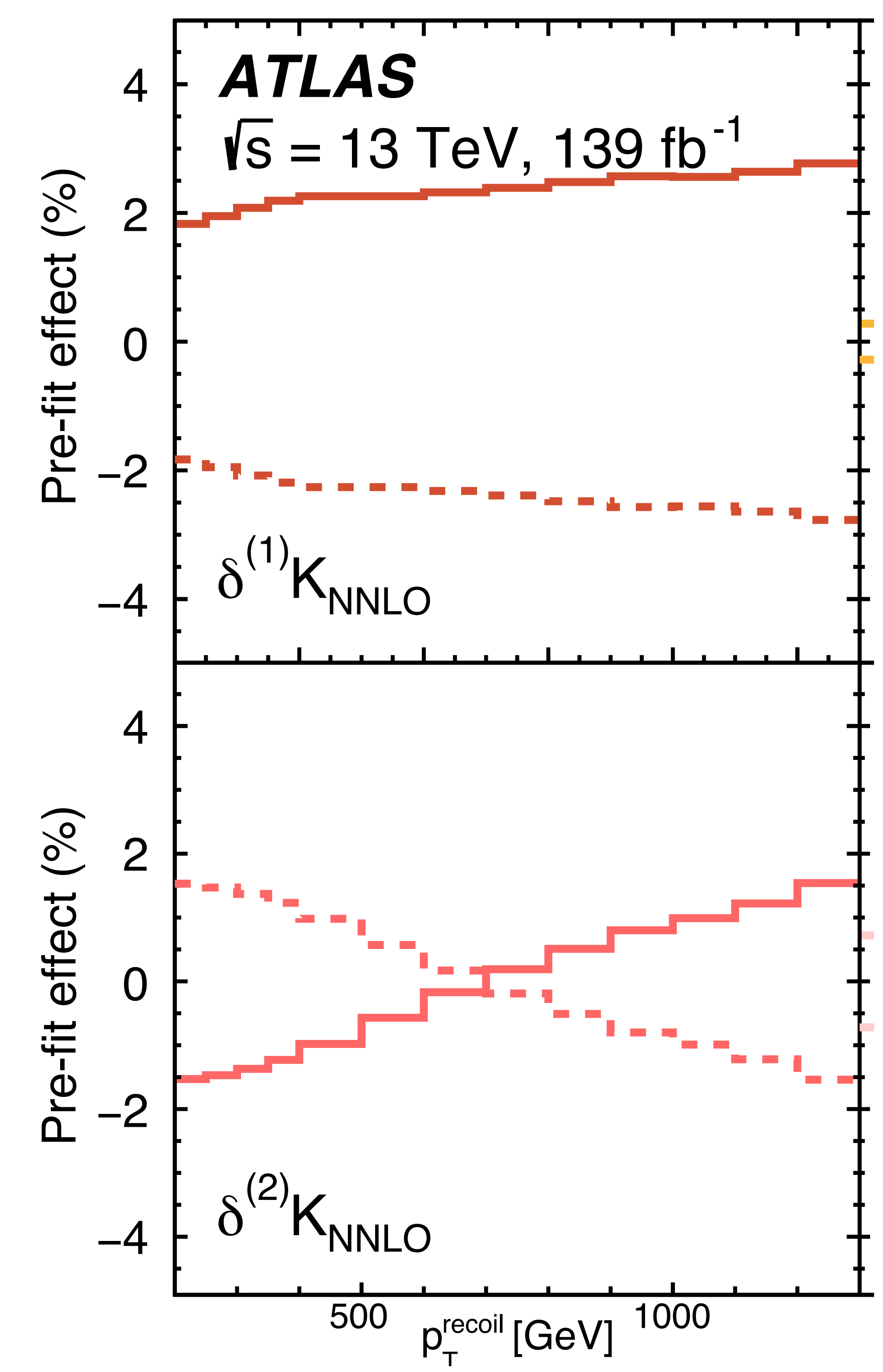


Conservative estimation of uncertainties due to scale variations

- **Standard**: envelope of the (μR, μF) = (1, 1), (2, 2), (0.5, 0.5), (2, 1), (1, 2), (1, 0.5), (0.5, 1) variations
  - Typically underestimates the shape uncertainty when correlated across different phase spaces
- **Lindert et al** ([1705.04664](https://arxiv.org/abs/1705.04664)): augment the envelope with an additional anti-correlated uncertainty
  - E.g. fix to +/- 100% of envelope at the edge of phase-space and flip the sign in the middle
- Combining the two uncertainties will allow for normalization and shape variations within the envelope
- This approach already used in a few ATLAS and CMS publications, but should be considered more often
  - Especially when correlating scale uncertainties across large energy scales or MVA bins

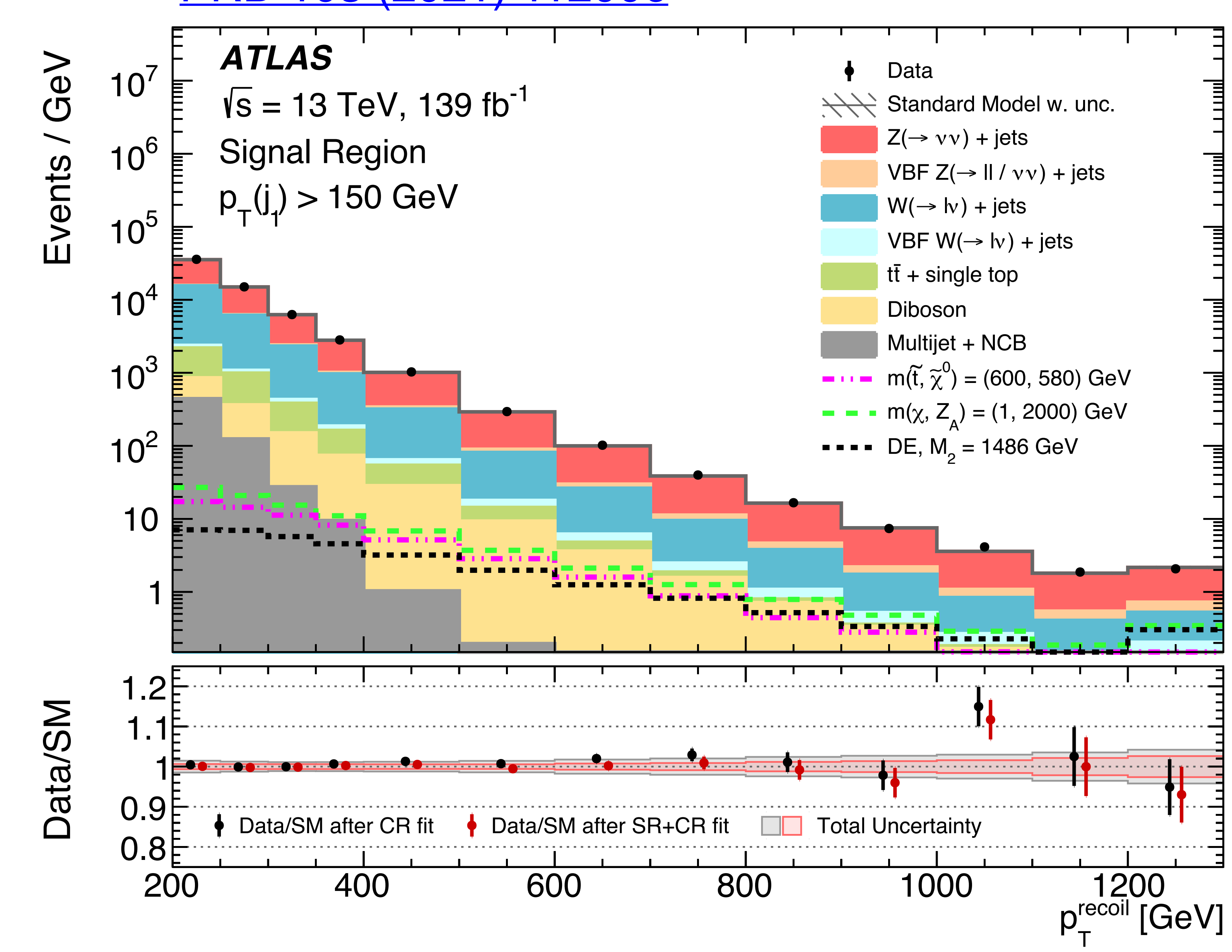
$$\omega_{\text{shape}}(p_T) = \tanh\left[\ln\left(\frac{p_T}{p_{T,0}}\right)\right] = \frac{p_T^2 - p_{T,0}^2}{p_T^2 + p_{T,0}^2}$$



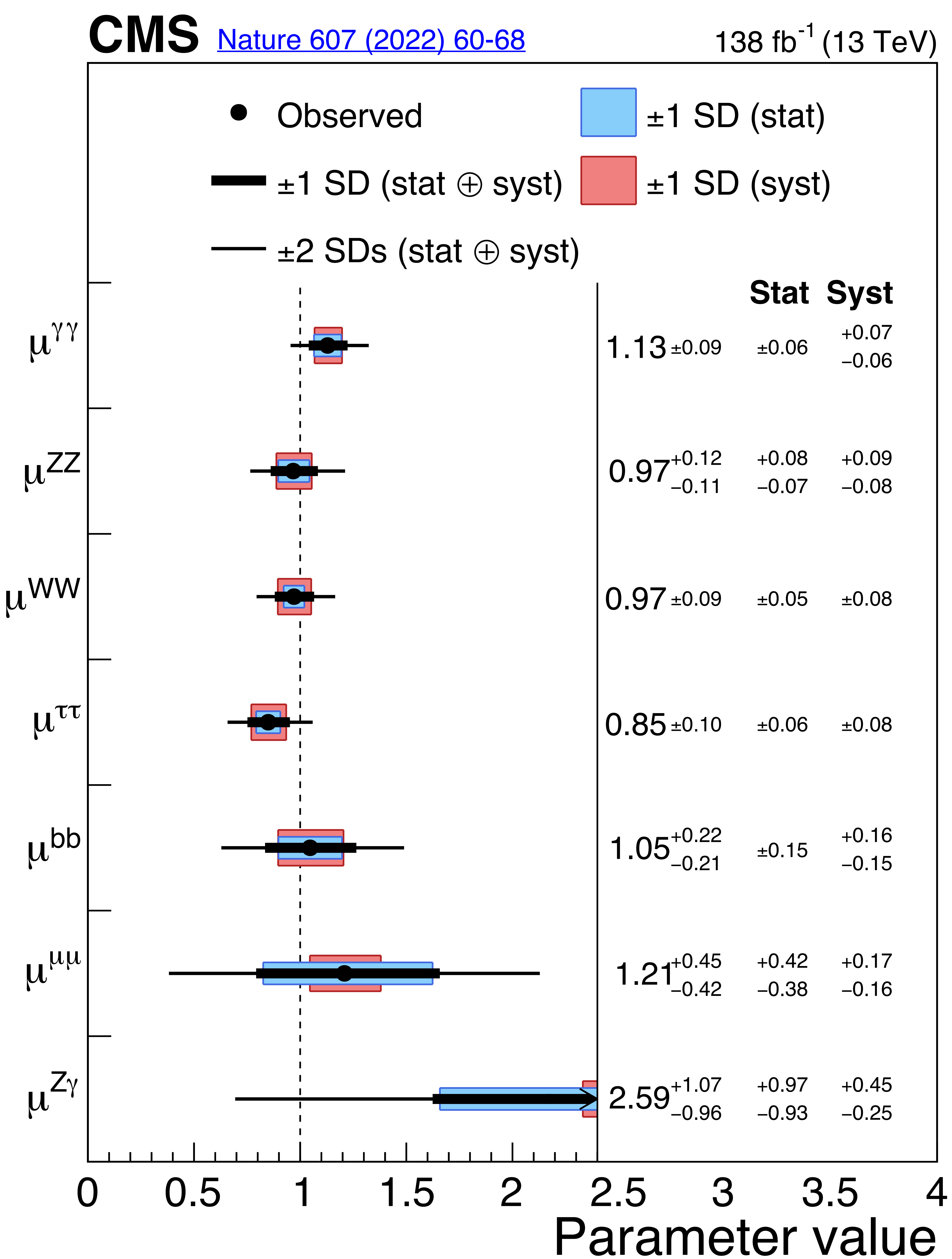


ATLAS MET + Jet search

[PRD 103 (2021) 112006](https://doi.org/10.1103/PhysRevD.103.112006)



QCD modeling limitations in (Di-)Higgs searches



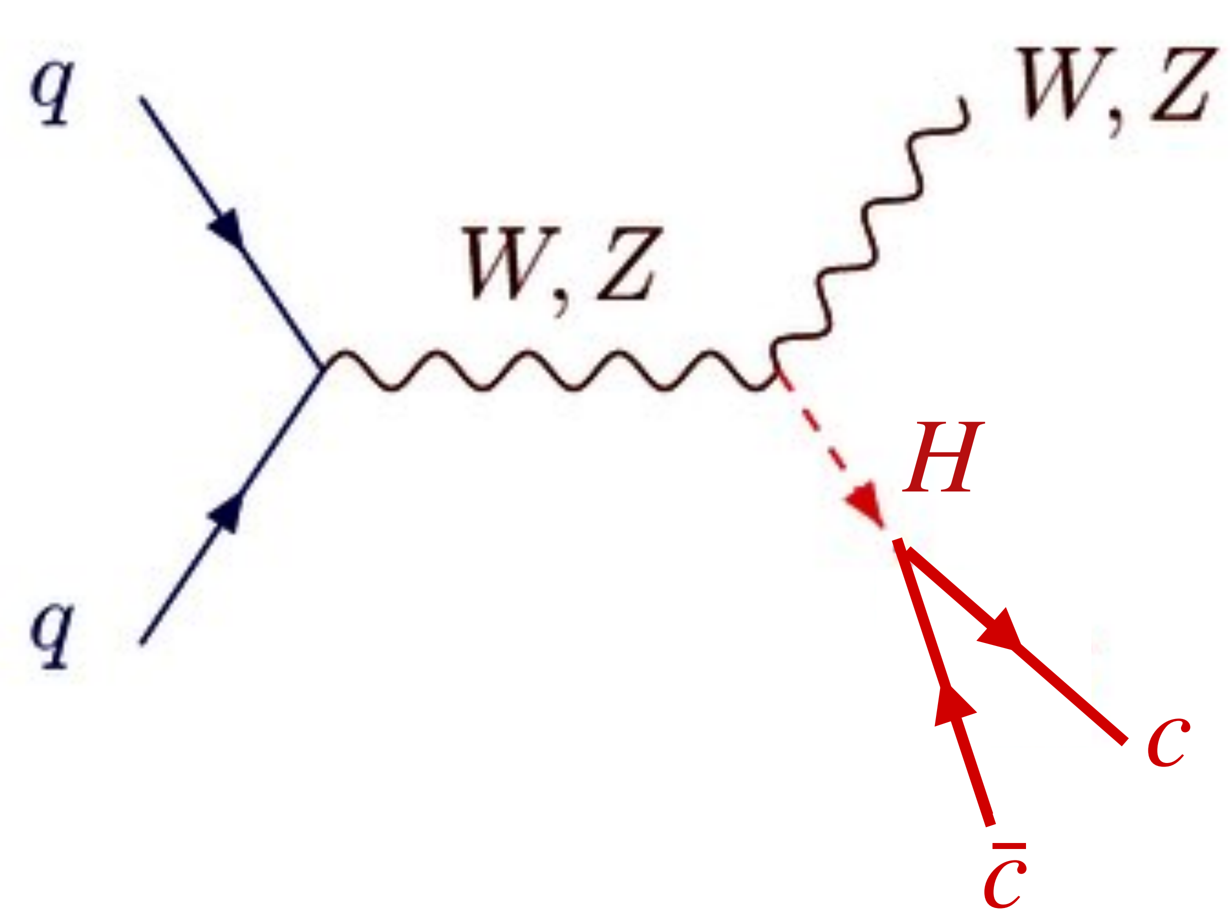
- For Higgs measurements reaching <10% precision, **signal QCD modeling becomes a large uncertainty**
- MC background modeling is most important for H→bb/cc, H→WW, and (surprisingly) H→μμ

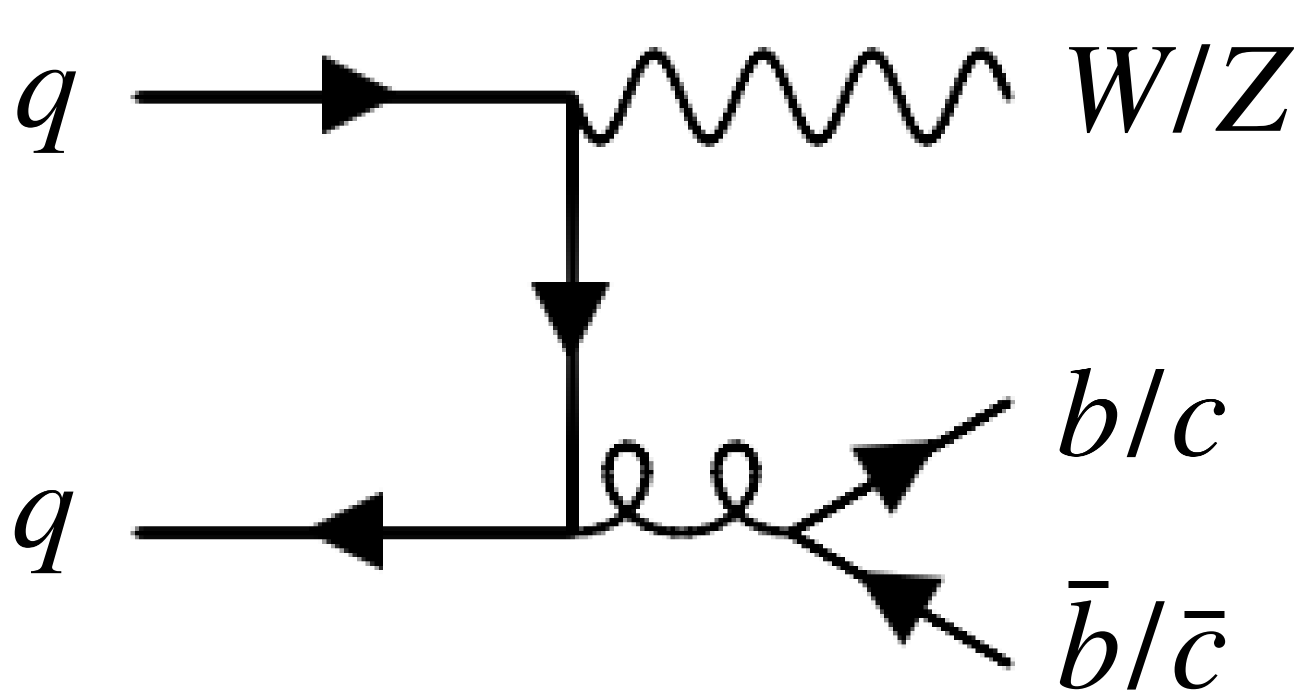
ATLAS H→γγ (data-driven background)

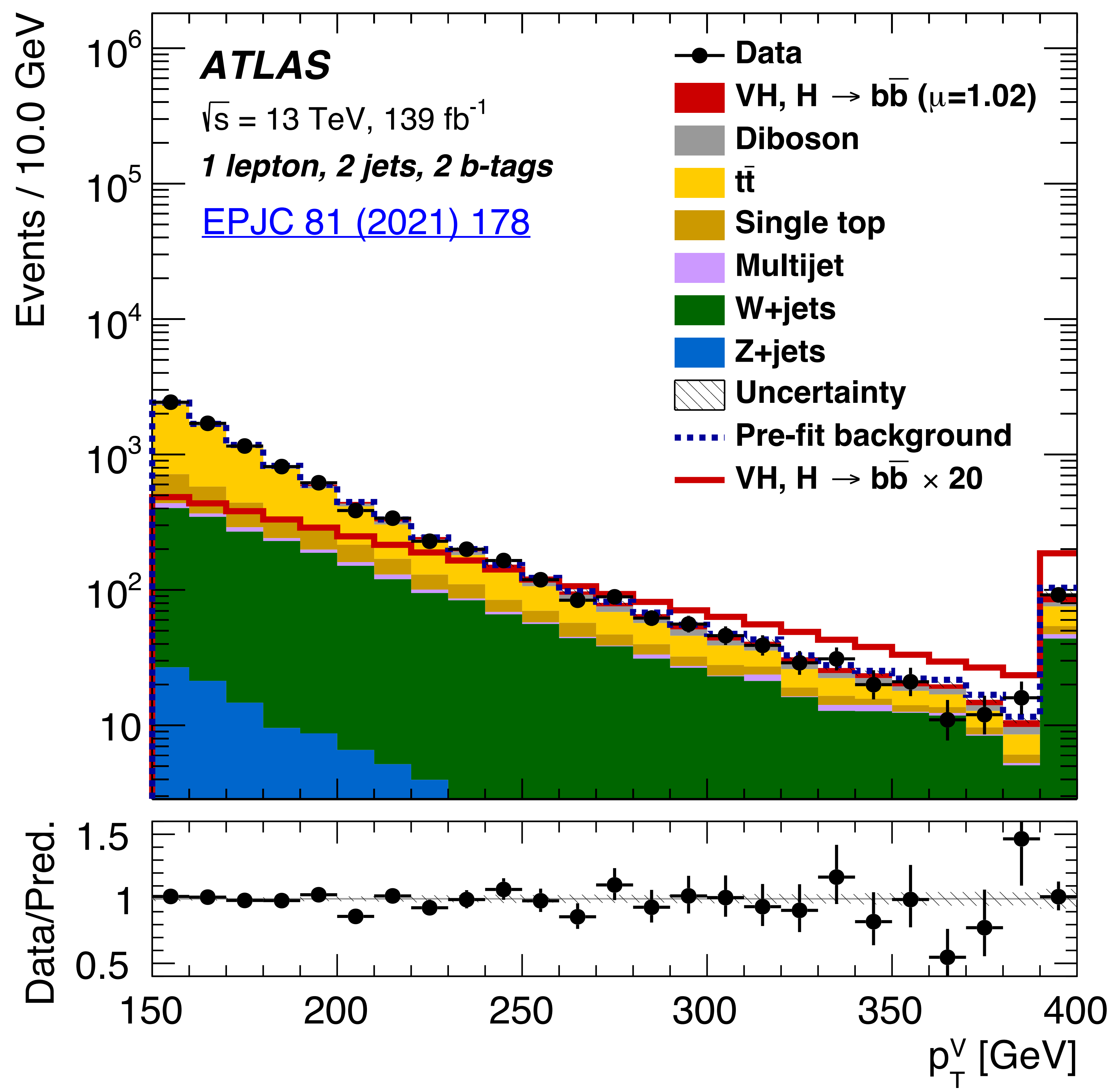
JHEP 07 (2023) 088

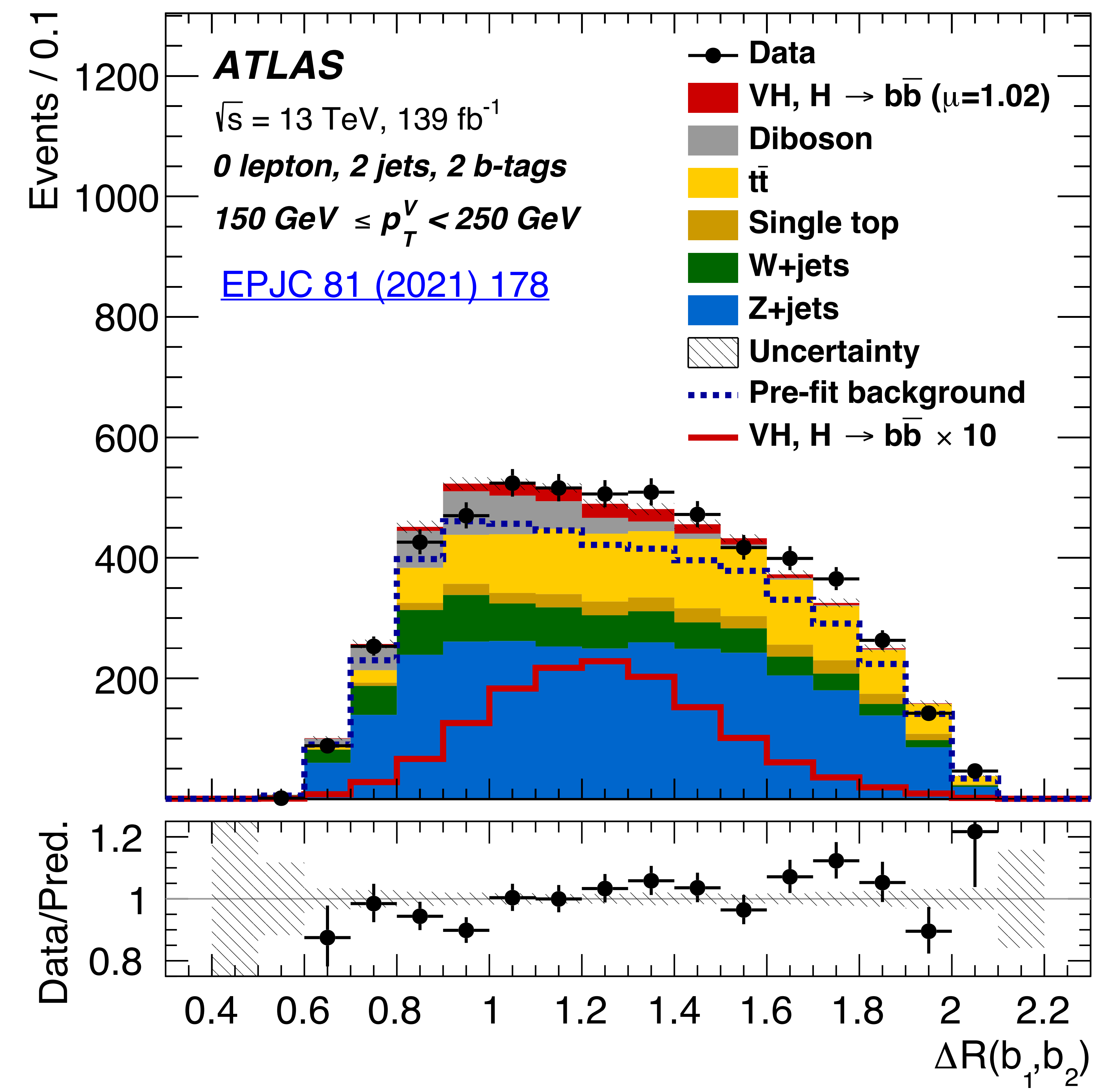
	ggF + $b\bar{b}$H	VBF	WH	ZH	$t\bar{t}H$
Uncertainty source	$\Delta\sigma$[%]	$\Delta\sigma$[%]	$\Delta\sigma$[%]	$\Delta\sigma$[%]	$\Delta\sigma$[%]
Theory uncertainties					
Higher-order QCD terms	±1.4	±4.1	±4.1	±12	±2.8
Underlying event and parton shower	±2.5	±16	±2.5	±4.0	±3.6
PDF and α_s	< ±1	±2.0	±1.4	±2.3	< ±1
Matrix element	< ±1	±3.2	< ±1	±1.2	±2.5
Heavy-flavour jet modelling in non-$t\bar{t}H$ processes	< ±1	< ±1	< ±1	< ±1	< ±1
Experimental uncertainties					
Photon energy resolution	±3.0	±3.0	±3.8	±4.8	±3.0
Photon efficiency	±2.7	±2.7	±3.3	±3.6	±2.9
Luminosity	±1.8	±2.0	±2.4	±2.7	±2.2
Pile-up	±1.4	±2.2	±2.0	±2.3	±1.4
Background modelling	±2.0	±4.6	±3.6	±7.2	±2.5
Photon energy scale	< ±1	< ±1	< ±1	±1.3	< ±1
Jet/E_T^{miss}	< ±1	±6.8	< ±1	±2.2	±3.5
Flavour tagging	< ±1	< ±1	< ±1	< ±1	±1.5
Leptons	< ±1	< ±1	< ±1	< ±1	< ±1
Higgs boson mass	< ±1	< ±1	< ±1	< ±1	< ±1

- So far H→bb/cc is most precisely constrained with the associated VH production mode
- Relies on flavor tagging to identify b- and c-jets and remove backgrounds from light jets
- Large remaining backgrounds from **V+bb/cc** with final state gluon splitting **g→bb/cc** and $t\bar{t}$+jets
- Relies on kinematic variables like $p_T(V)$, **Δ**R(bb), m(bb) to increase signal / background separation
  - Accurate modeling in MC is essential to allow extrapolations of the fit model from CRs to SRs

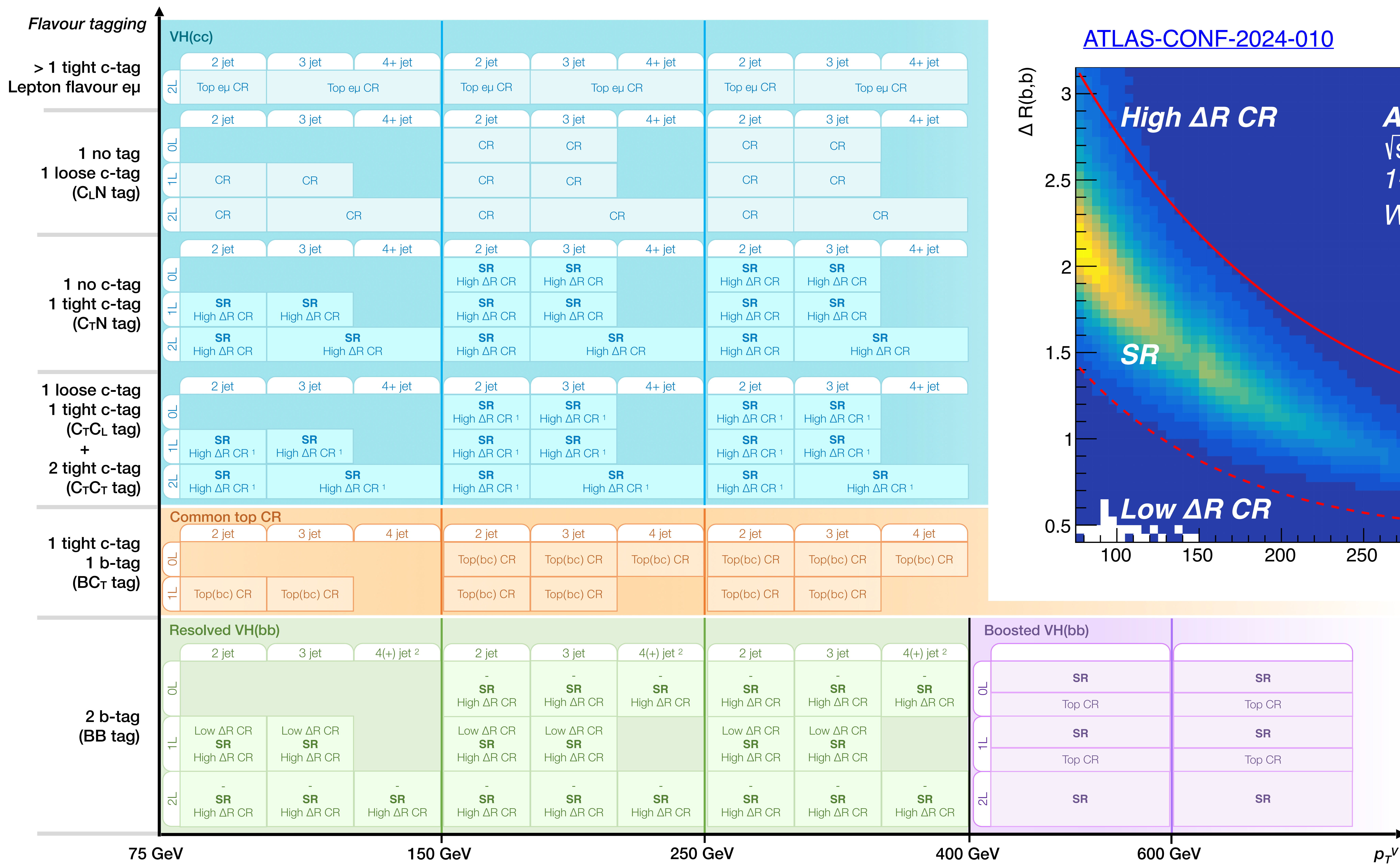




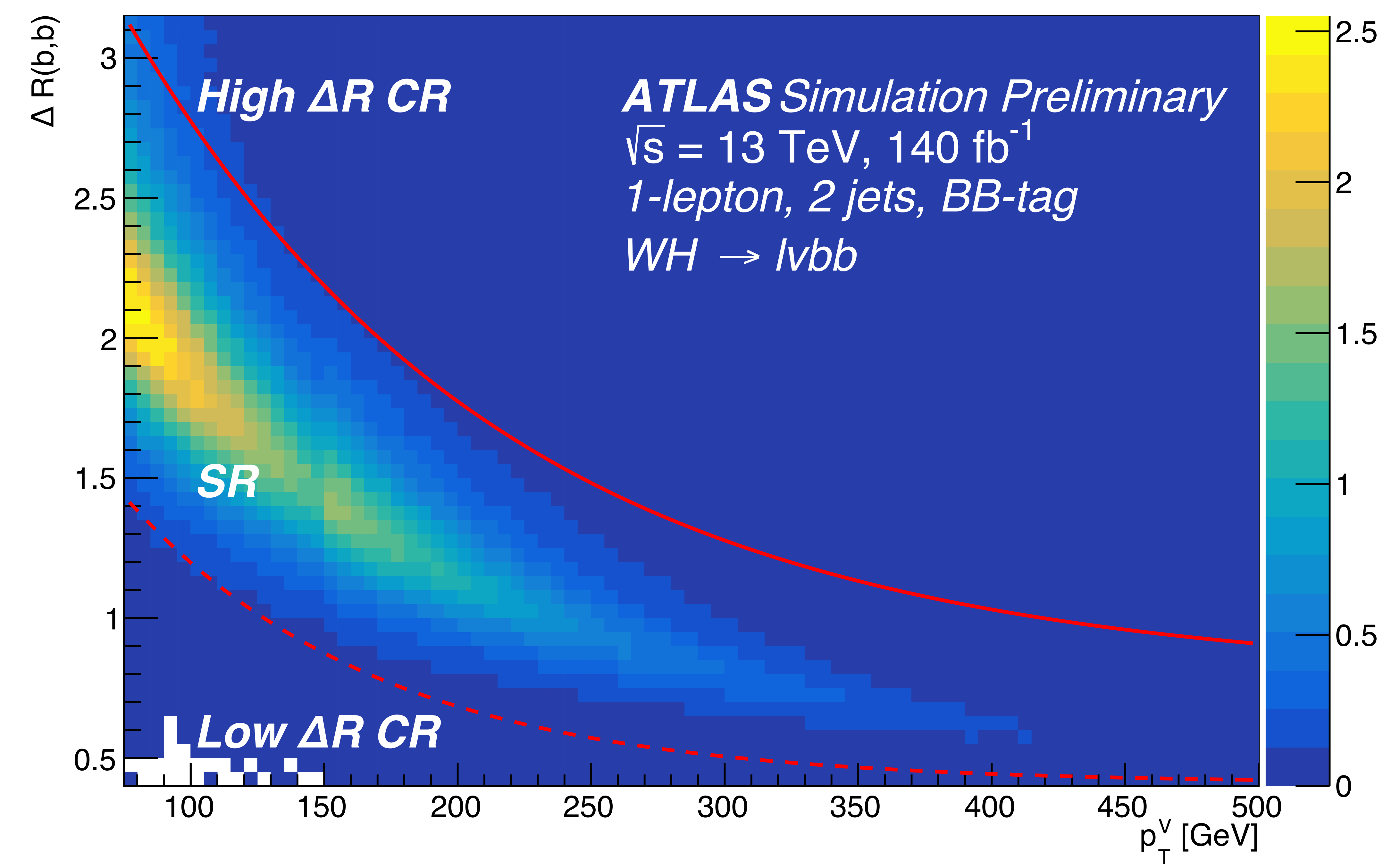




- Extract the VH(bb) and VH(cc) signals simultaneously— more details by Alexei Raspereza (talk)
- **Likelihood fit**: 97 CRs + 59 SRs with 42 normalization factors for V+jets and 23 for $t\bar{t} + Wt$
  - Additional ~100 nuisance parameters related to background modeling



ATLAS-CONF-2024-010



Impact of background theory uncertainties in H→bb/cc

ATLAS

$\mu_{VH}^{bb} = 0.91 \pm 0.10\,(\text{stat.}) \pm 0.12\,(\text{syst.})$

$\mu_{VH}^{cc} = 1.0 \pm 4.0\,(\text{stat.}) \pm 3.5\,(\text{syst.})$

Source of uncertainty		$VH, H \to b\bar{b}$	$VH, H \to c\bar{c}$
Total		0.151	5.29
Statistical		0.097	3.94
Systematic		0.116	3.53
Statistical uncertainties			
Data statistical		0.089	3.70
$t\bar{t}$ $e\mu$ control region		0.009	0.06
Background floating normalisations		0.034	1.23
Other VH floating normalisation		0.007	0.24
Simulation samples size		0.023	1.61
Experimental uncertainties			
Jets		0.028	1.00
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.009	0.24
Leptons		0.004	0.23
b-tagging	b-jets	0.020	0.30
	c-jets	0.013	0.73
	light-flavour jets	0.006	0.67
Pile-up		0.009	0.24
Luminosity		0.006	0.08
Theoretical and modelling uncertainties			
Signal		0.073	0.56
Z + jets		0.039	1.76
W + jets		0.055	1.41
$t\bar{t}$ and Wt		0.018	1.03
Single top quark (s-, t-ch.)		0.010	0.15
Diboson		0.032	0.51
Multi-jet		0.006	0.57

ATLAS-CONF-2024-010

- Background modeling related uncertainties are among the largest in the ATLAS and CMS H→bb/cc searches
- MC statistical uncertainty also poses an issue due the computational complexity in generating the multi-leg NLO samples
- At HL-LHC it will be extremely challenging to get background predictions with systematic uncertainty matching the data statistics

CMS $\mu_{VH}^{bb} = 1.15 \pm 0.21$

	$\Delta\mu$	
Background (theory)	+0.043	−0.043
Signal (theory)	+0.088	−0.059
MC sample size	+0.078	−0.078
Simulation modeling	+0.059	−0.059
b tagging	+0.050	−0.046
Jet energy resolution	+0.036	−0.028
Int. luminosity	+0.032	−0.027
Jet energy scale	+0.025	−0.025
Lepton ident.	+0.008	−0.007
Trigger ($\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$)	+0.002	−0.001

PRD 109 (2024) 092011

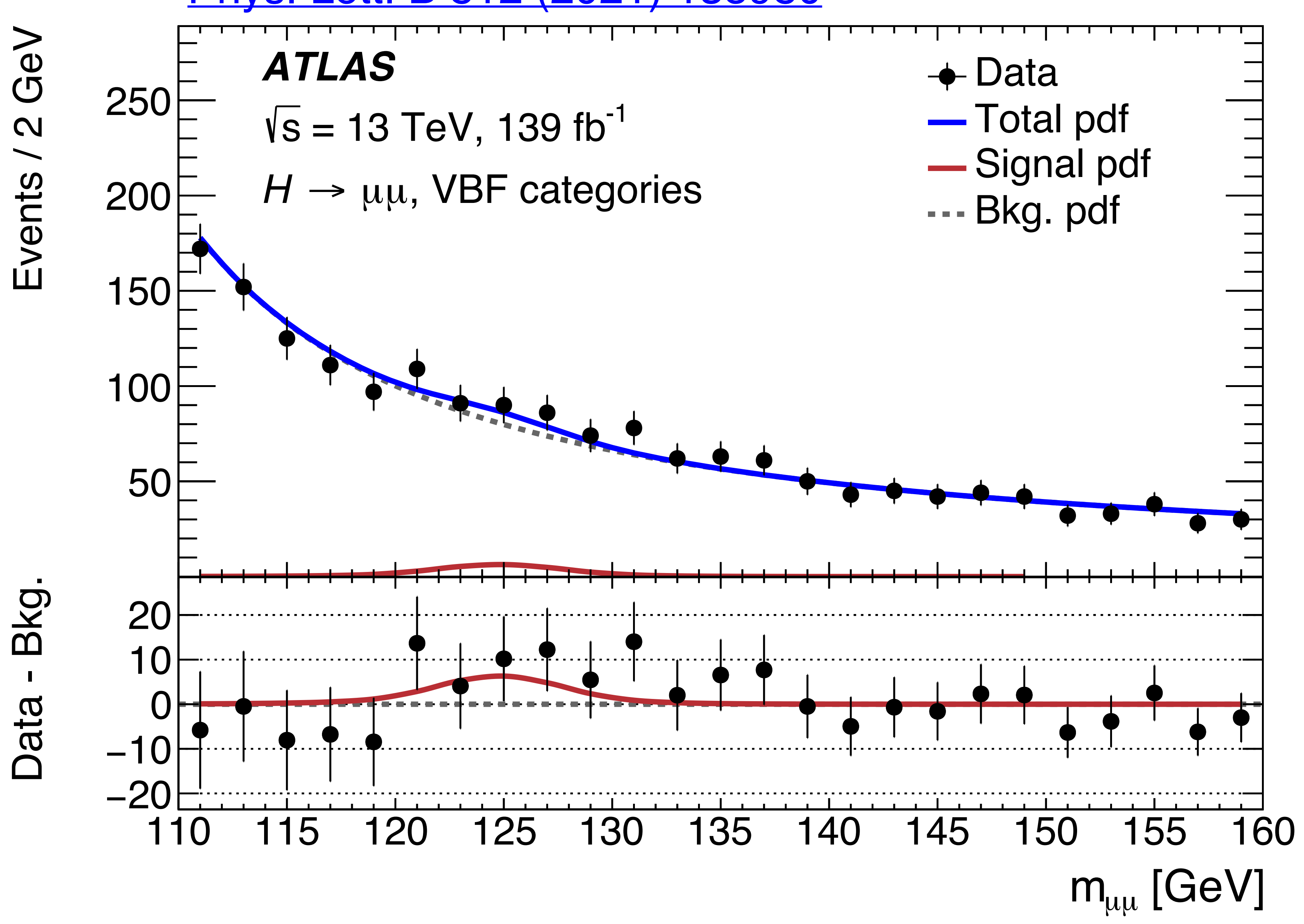
CMS $\mu_{VH}^{cc} = 7.7 \pm 3.7$

Uncertainty source	$\Delta\mu / (\Delta\mu)_{\mathrm{tot}}$
Statistical	85%
Background normalizations	37%
Experimental	48%
Sizes of the simulated samples	37%
c jet identification efficiencies	23%
Jet energy scale and resolution	15%
Simulation modeling	11%
Integrated luminosity	6%
Lepton identification efficiencies	4%
Theory	22%
Backgrounds	17%
Signal	15%

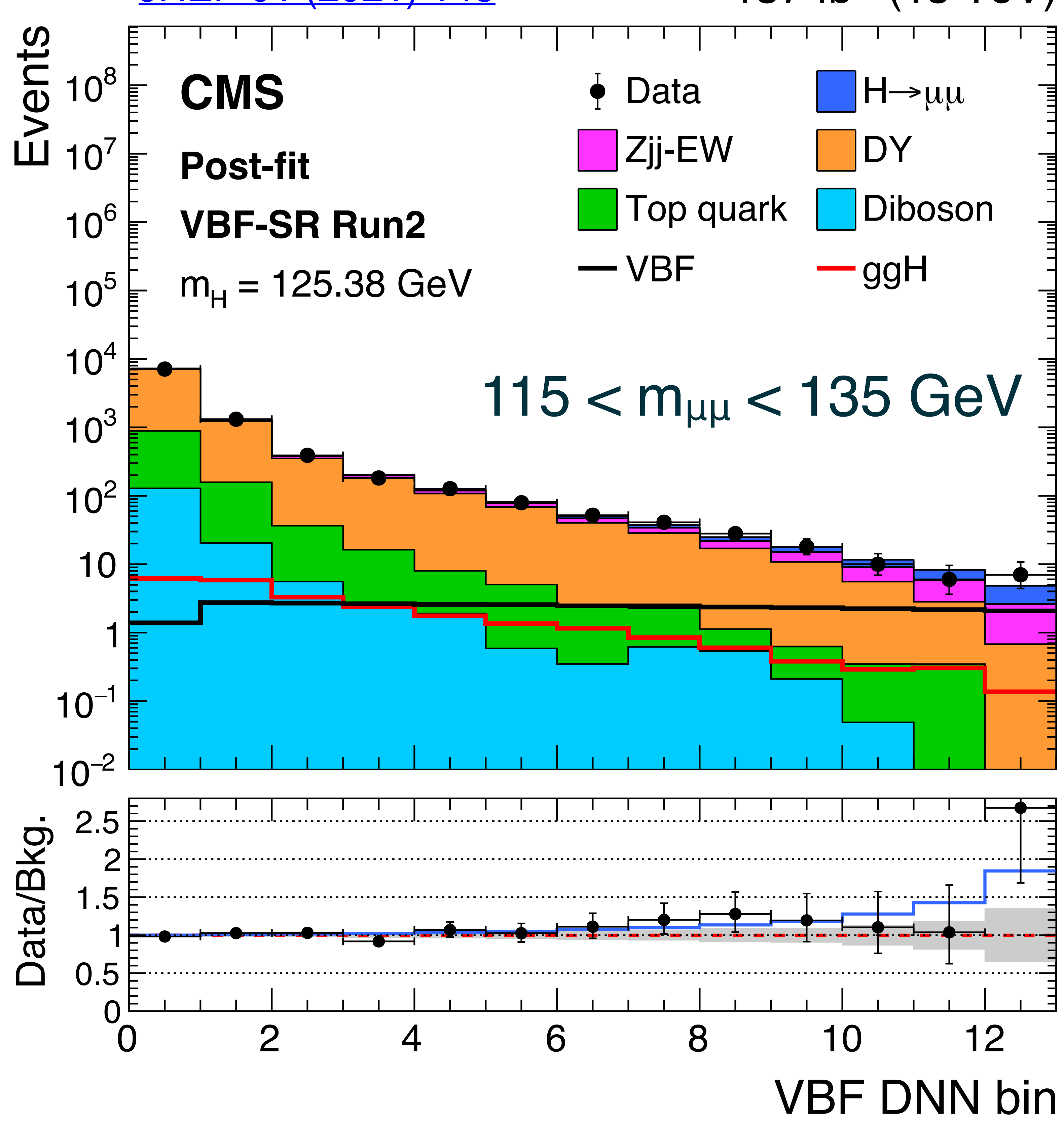
PRL 131 (2023) 061801

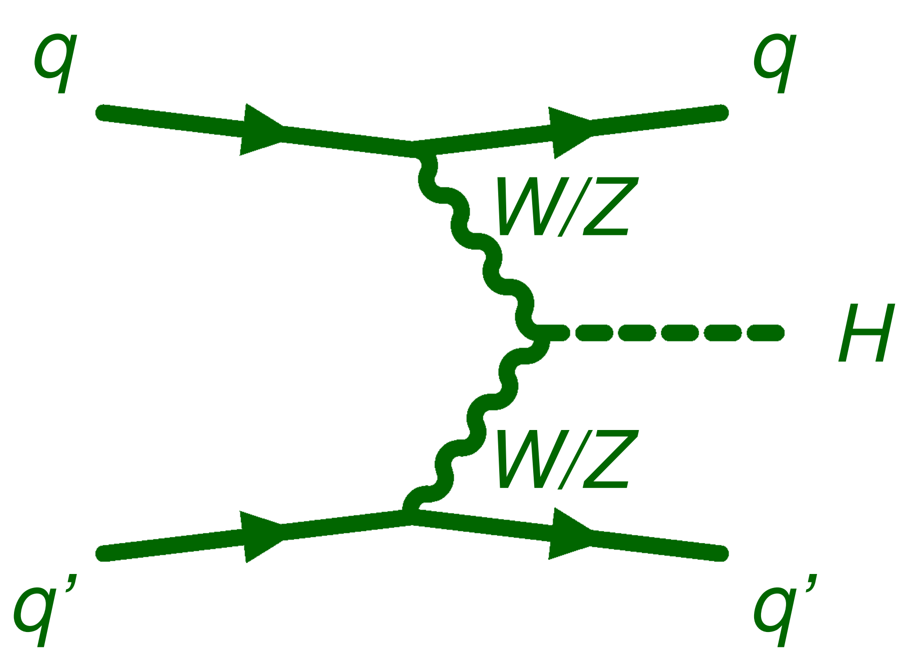
- So far ATLAS had fully data-driven background modeling approach for the main channels (ggF and VBF)
- CMS gained extra sensitivity by using MC modeling in the VBF category and achieved 3σ evidence
- Accurate MC prediction for VBF Z necessary to allow extrapolations from $m_{\mu\mu}$ side-bands into the peak region and to allow fitting a kinematic multi-variate DNN discriminant
- Potential to reach 5σ sensitivity by the end of Run 3— need to be very confident in the background model

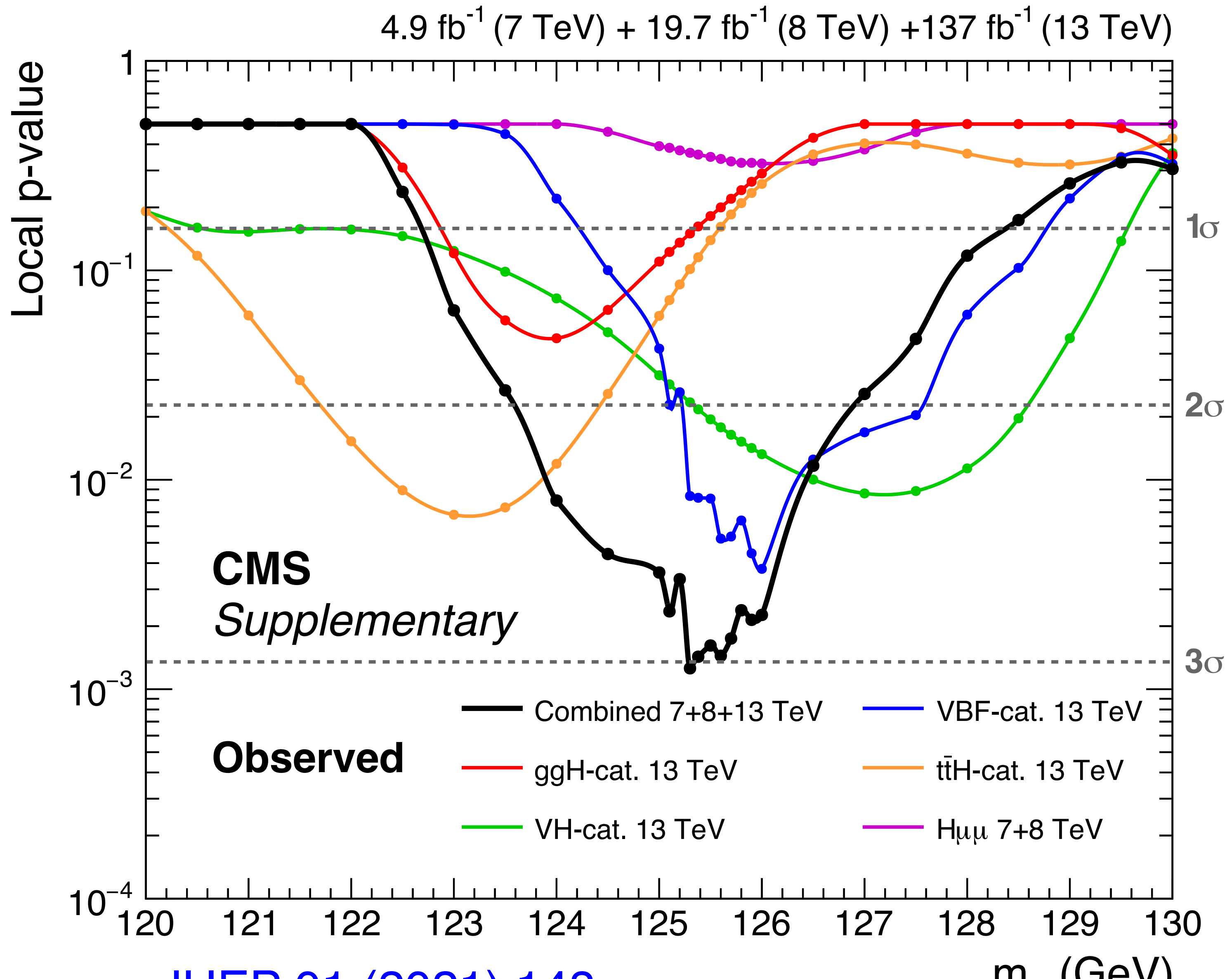
Phys. Lett. B 812 (2021) 135980



JHEP 01 (2021) 148

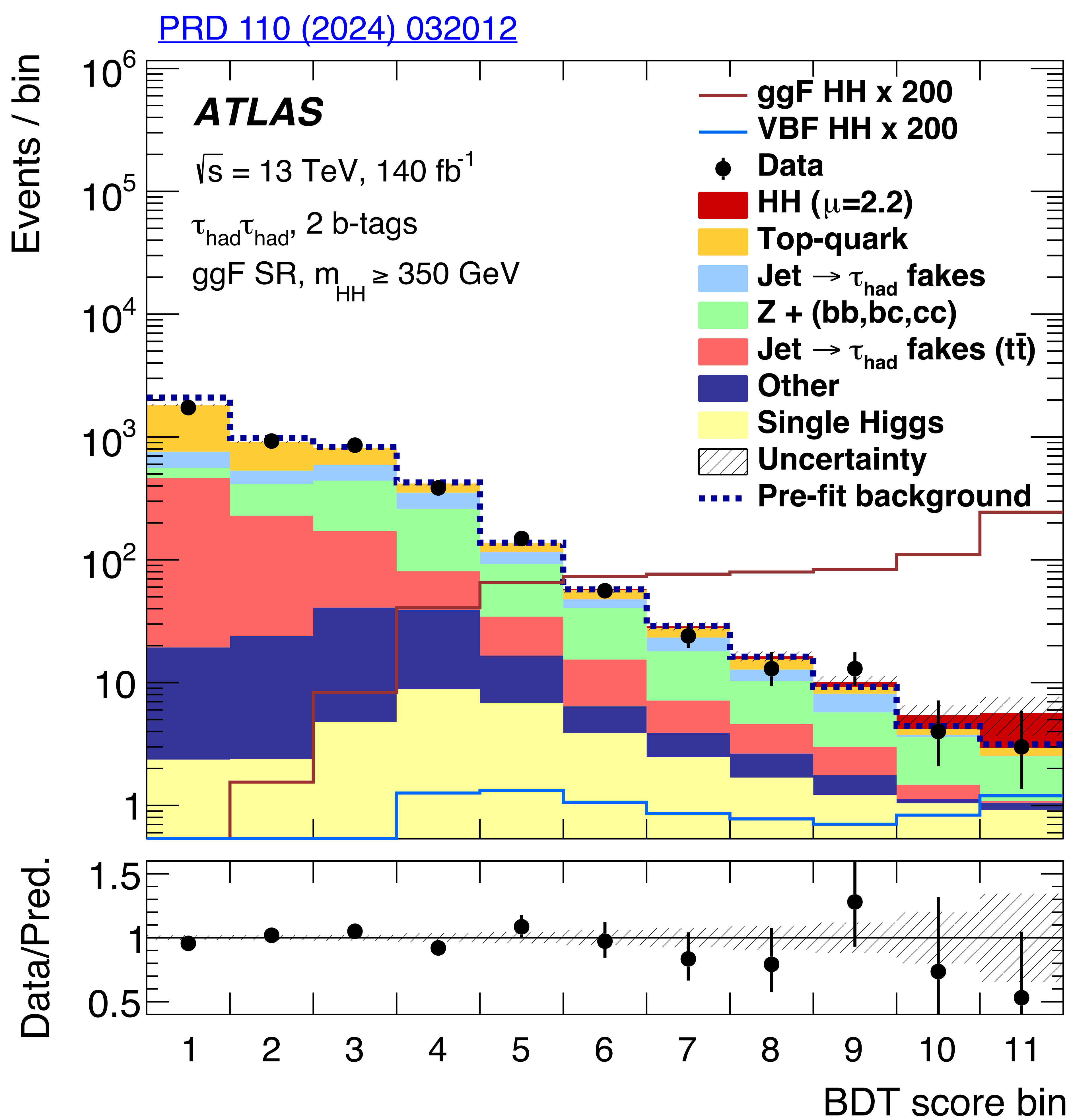


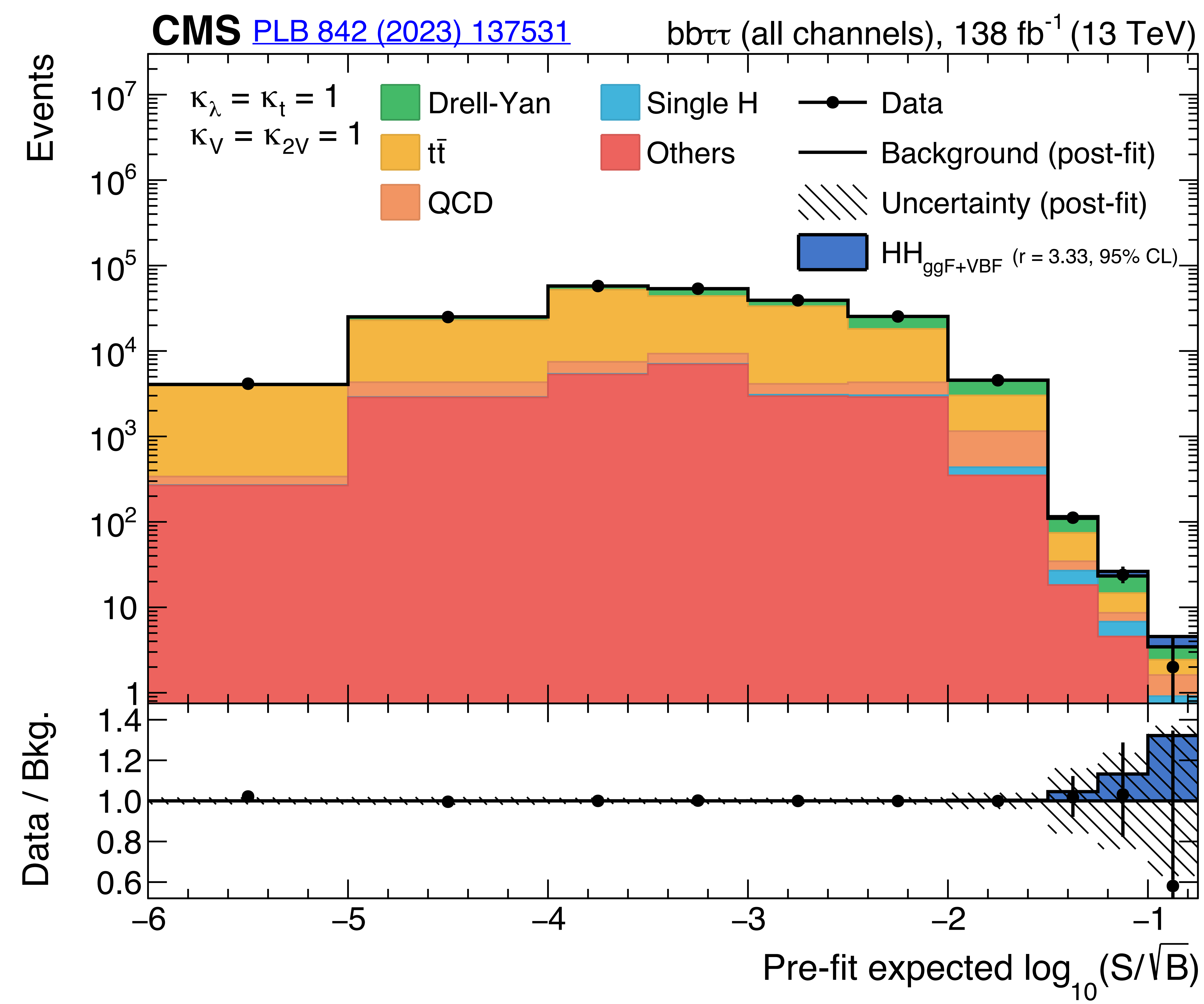




JHEP 01 (2021) 148

- Among the main channels (4b, bbγγ, bb$\tau\tau$); bb$\tau\tau$ relies the most on MC-based background modeling
- Currently dominated by statistical uncertainties, but background modeling will become more and more important as we take more data at the LHC
- **Main backgrounds**: $Z \to \tau\tau + g(\to bb)$, $t\bar{t}$, Wt (interference effects have large impact..)

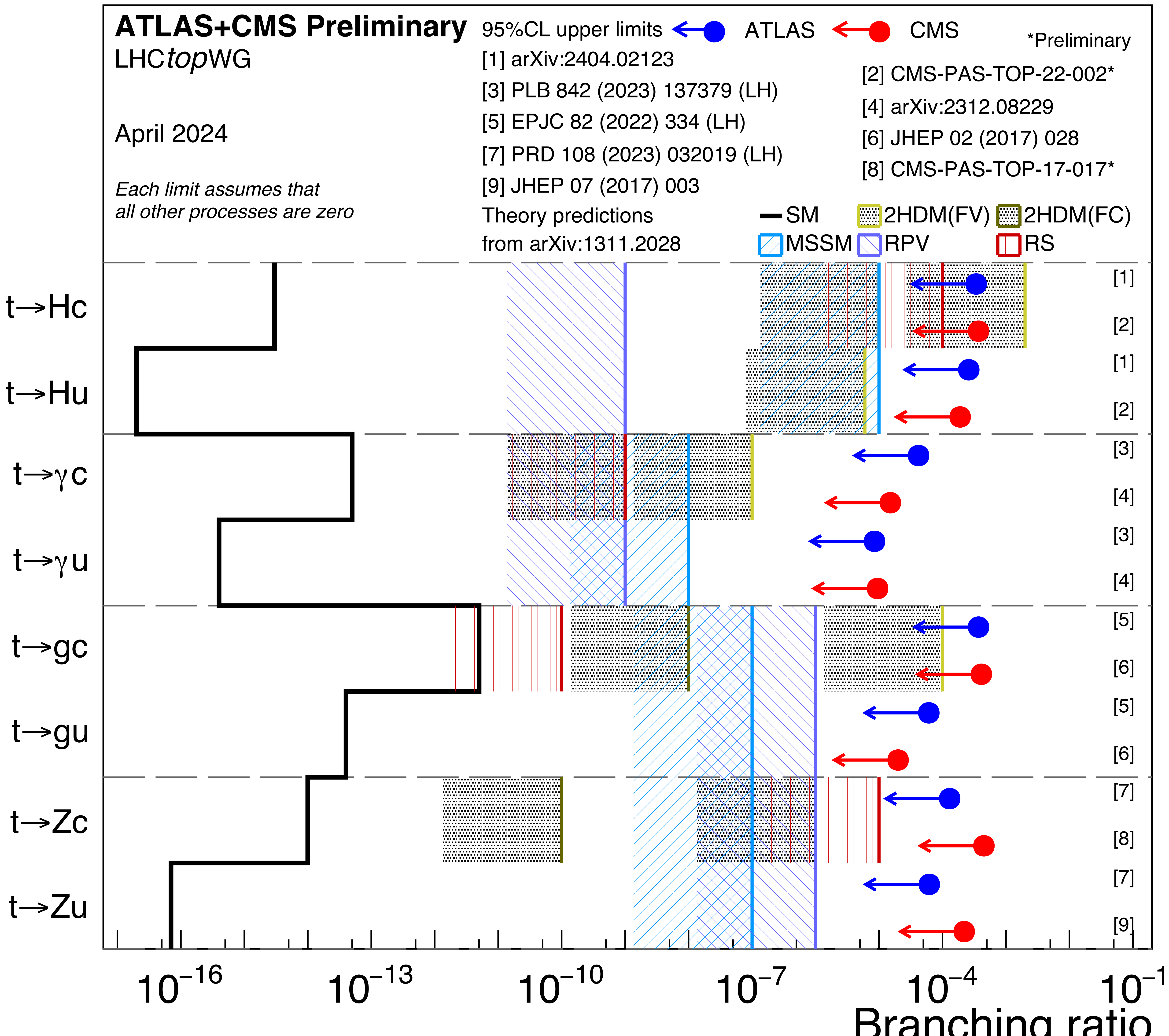


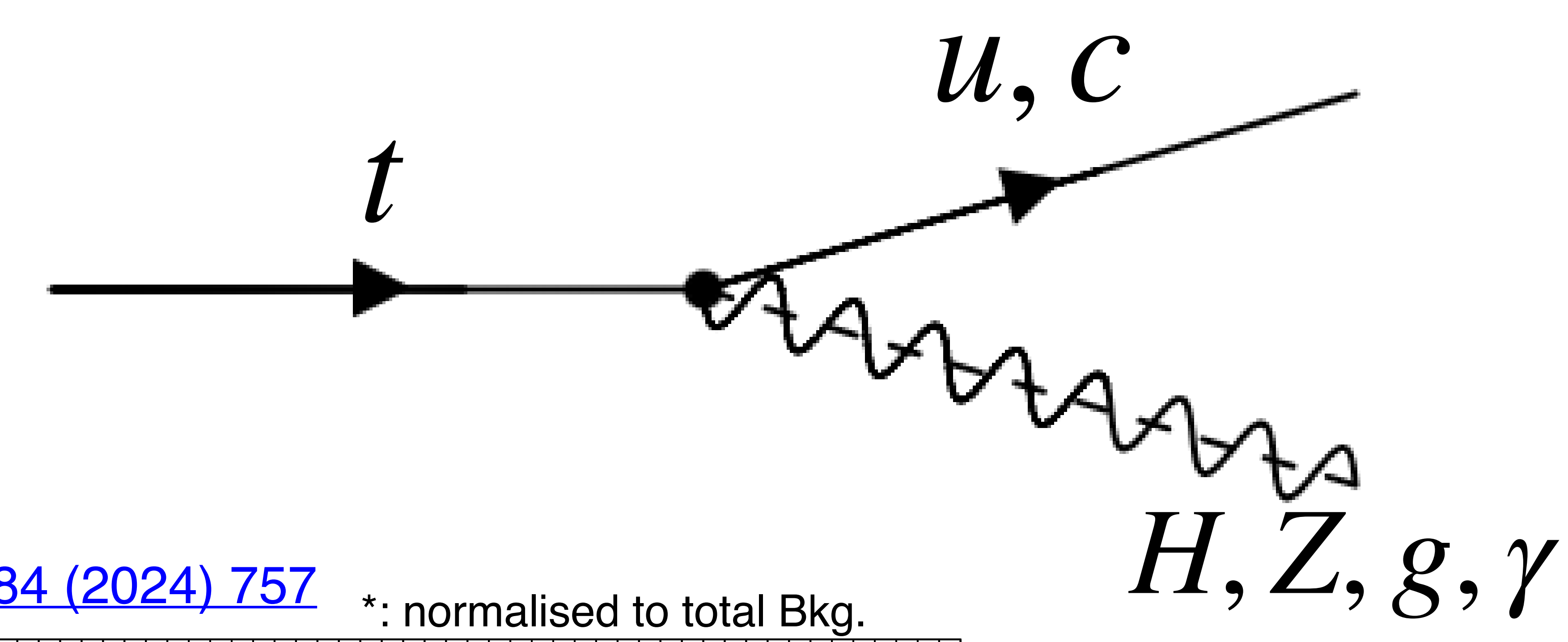


Background modeling limitations BSM searches

- Searching for rare Flavor Changing Neutral Current, heavily suppressed in the SM
- Most FCNC searches rely on MC modeling of the main backgrounds from EWK + QCD (e.g. $t\bar{t}V$)

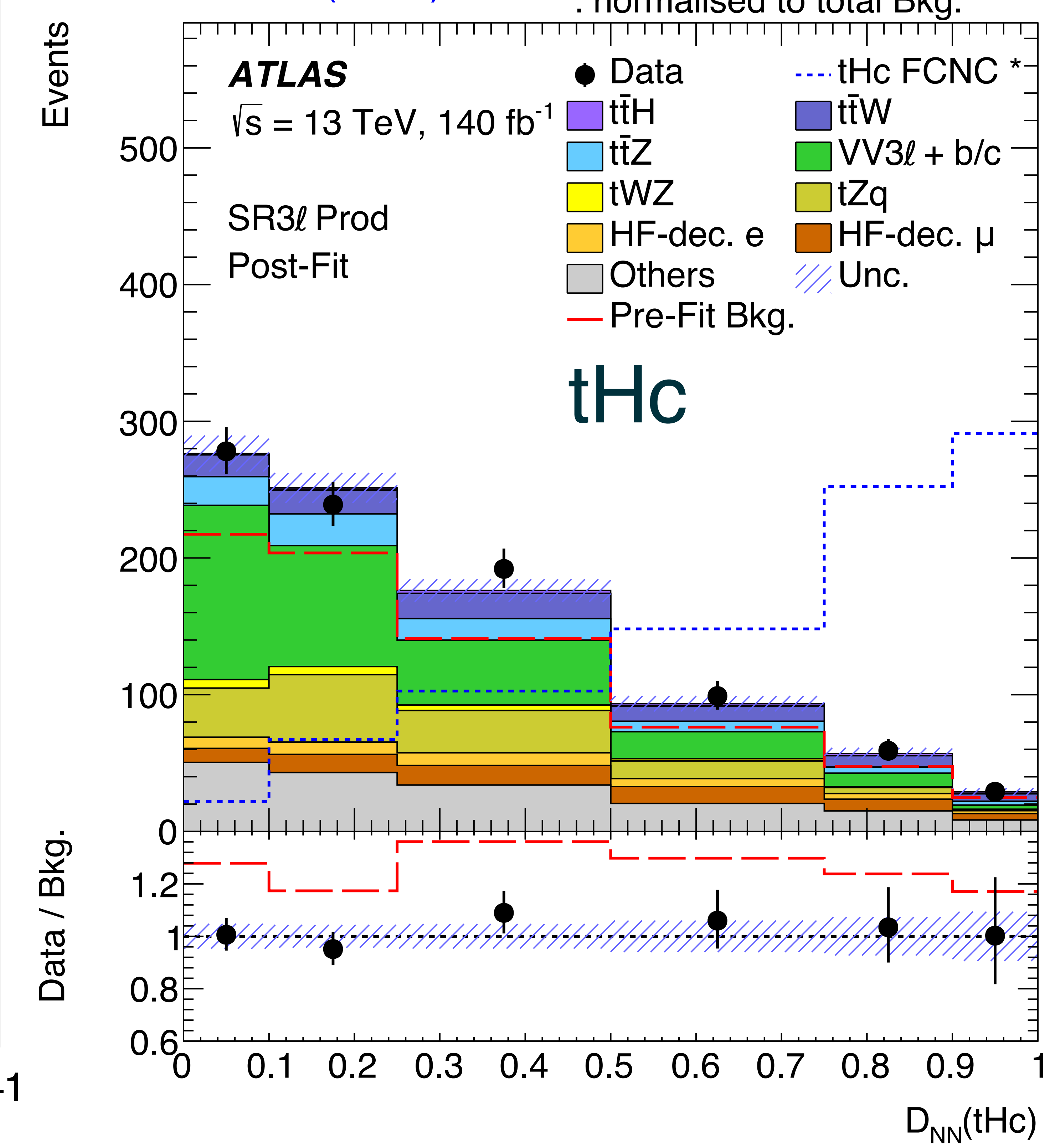
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots





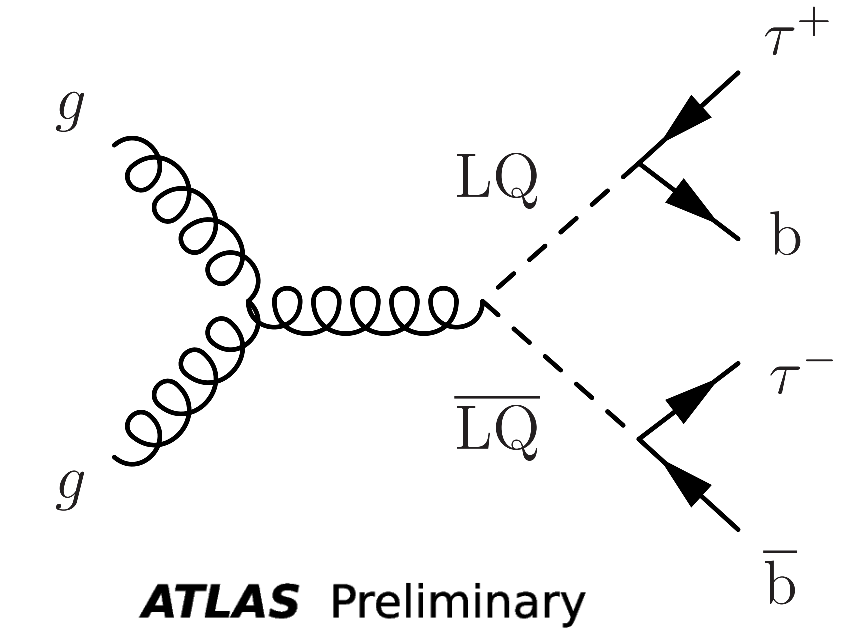
EPJC 84 (2024) 757

*: normalised to total Bkg.

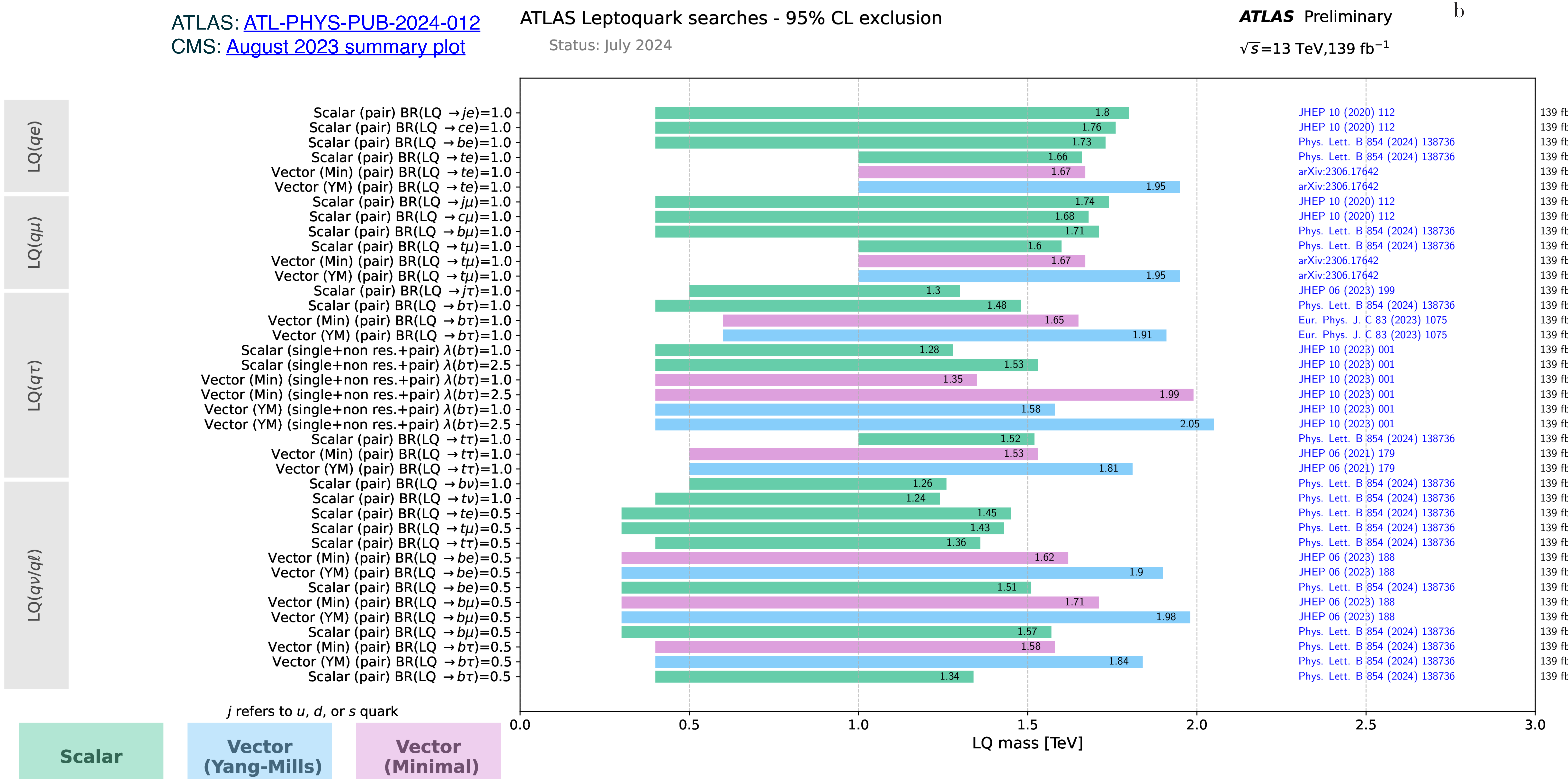


Leptoquark (LQ) searches

- Pair produced LQs generally give a signature of ℓℓ+qq (e.g. $\tau\tau$+bb, μμ+bb, …)
- Large Standard Model backgrounds from $Z \rightarrow \ell\ell + \text{jets}$, and $t\bar{t}$

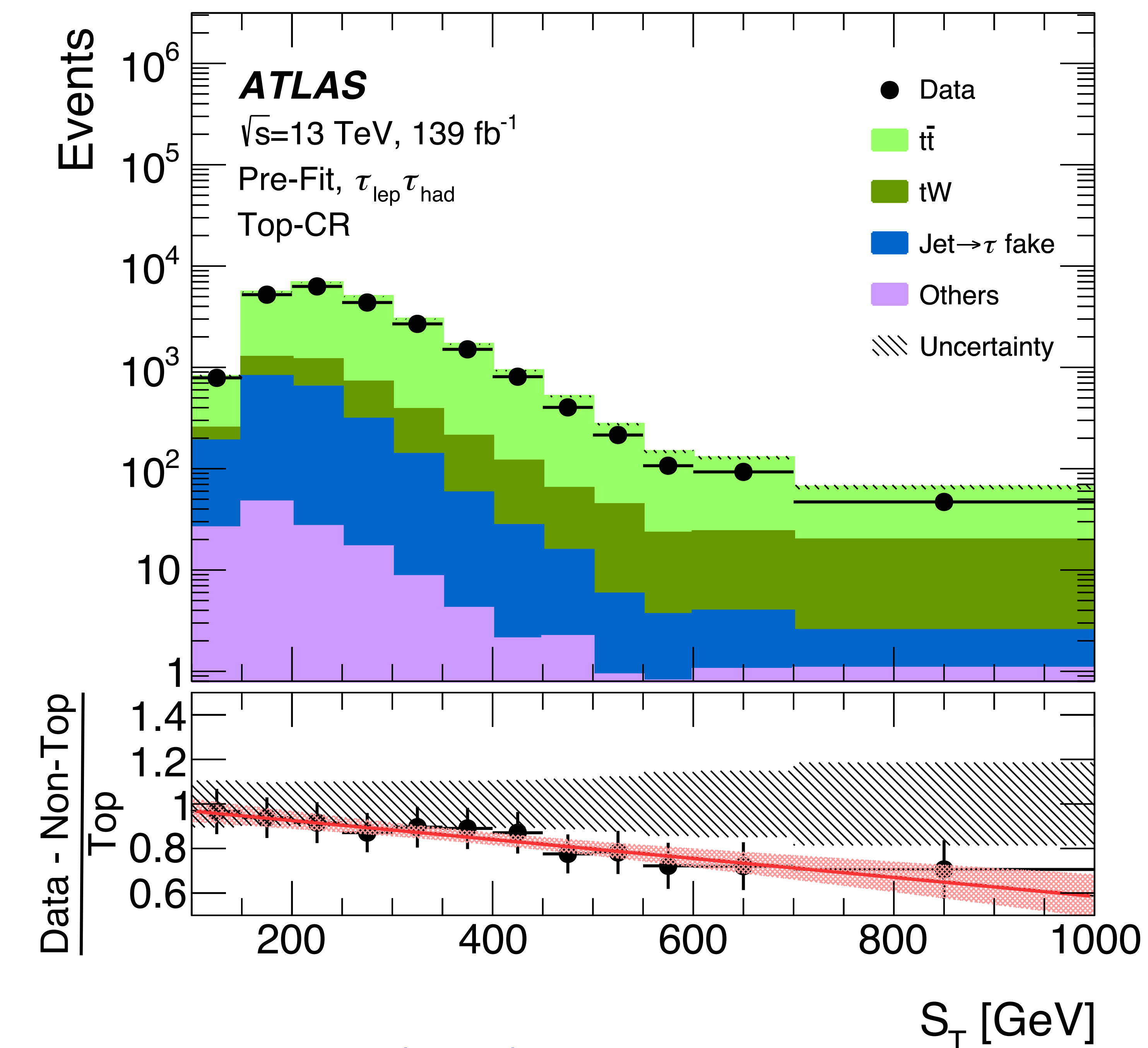


ATLAS: ATL-PHYS-PUB-2024-012
CMS: August 2023 summary plot

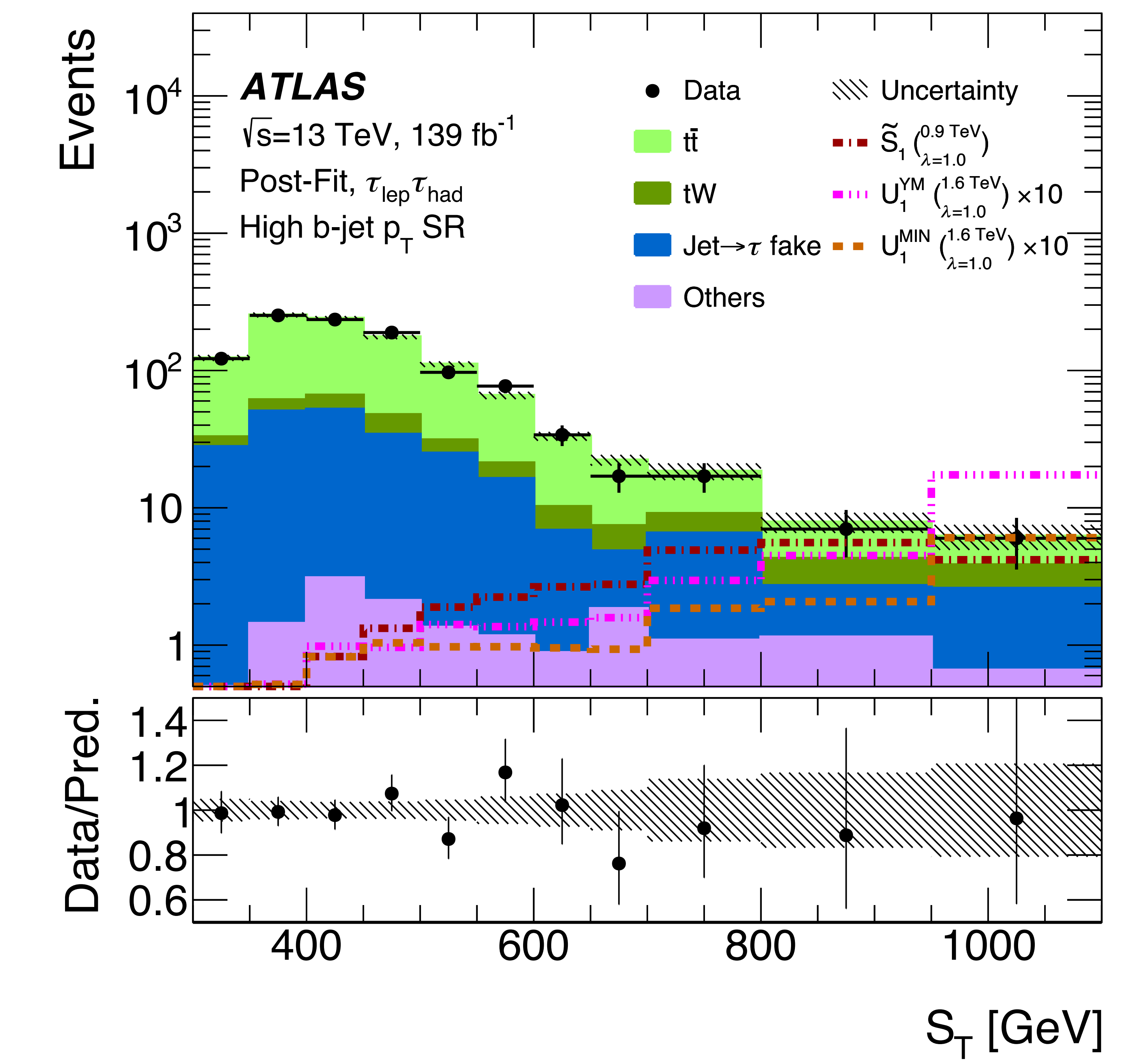


Leptoquark (LQ) searches example

- The LQ→τb analysis has large irreducible backgrounds form $t\bar{t}$
- Accurate modeling of top-quark $\vec{p}_t$ needed up to large momenta
- Due to large statistics the data became sensitive to NNLO corrections in top-quark $\vec{p}_t$— data-driven corrections implemented in this case
- Dominated by statistical uncertainty; up to 20% sensitivity loss from background modeling



JHEP 10 (2023) 001



Relative increase in the expected 95% CL upper limits on the cross-section times branching ratio relative to the statistical error only JHEP 10 (2023) 001

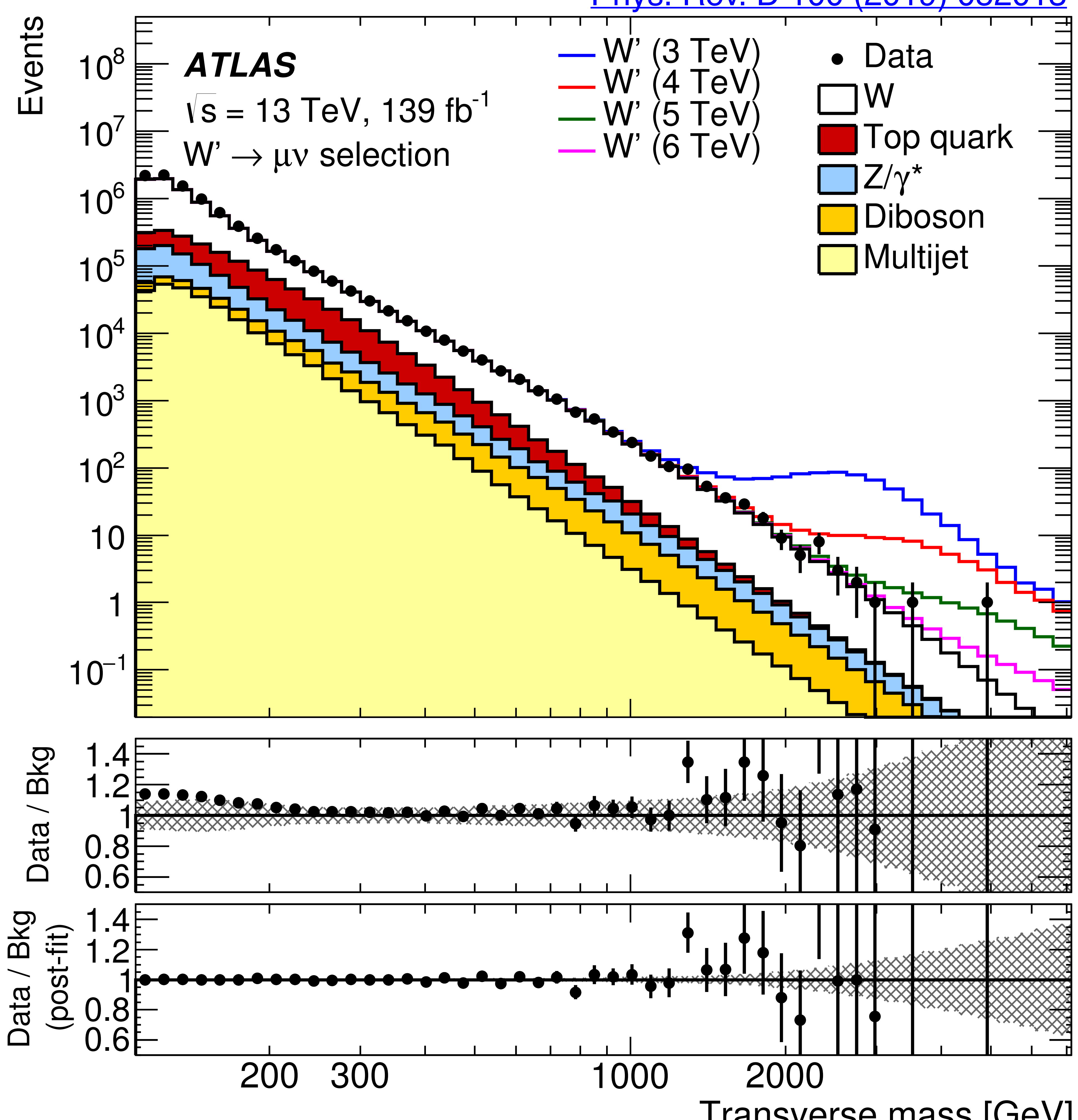
Source / m_{LQ}	$\lambda = 1.0$ 0.9 TeV	$\lambda = 1.0$ 1.6 TeV	$\lambda = 1.7$ 0.9 TeV	$\lambda = 1.7$ 1.6 TeV	$\lambda = 2.5$ 0.9 TeV	$\lambda = 2.5$ 1.6 TeV
Top background modelling	0.13	0.06	0.14	0.10	0.15	0.14
τ-lepton reconstruction/identification	0.07	0.06	0.07	0.07	0.08	0.07
τ-lepton energy scale	0.06	0.04	0.06	0.06	0.07	0.06
Flavour tagging	0.04	0.05	0.04	0.04	0.04	0.04
Signal acceptance	0.02	0.04	0.02	0.04	0.05	0.05
Others	0.05	0.04	0.05	0.04	0.05	0.05
MC statistics	0.09	0.08	0.09	0.08	0.09	0.08
Total	0.26	0.20	0.28	0.25	0.36	0.33

PDF uncertainties

Impact of PDF uncertainty on high mass resonance searches

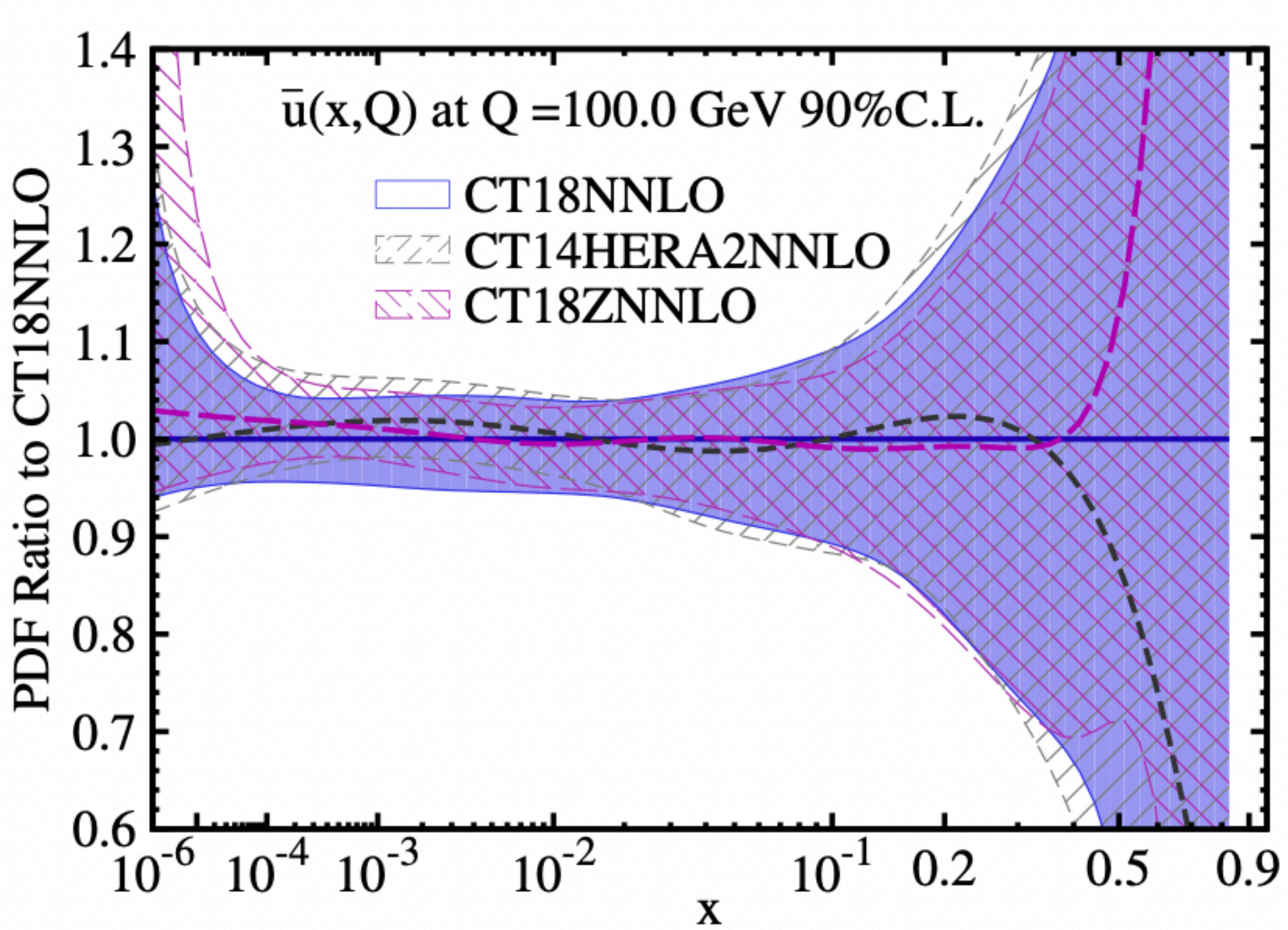
- PDF uncertainty is most relevant for searches at very large scales with large Drell—Yan backgrounds
  - These analyses are sensitive to PDFs at large Bjorken-x (e.g. x > 0.2) with large uncertainties!
- Current lower limits on Z' and W' mass are around 5-6 TeV; at HL-LHC it will be difficult to substantially improve Z' and W' sensitivity without reducing the Drell—Yan PDF uncertainty at high-x

Phys. Rev. D 100 (2019) 052013



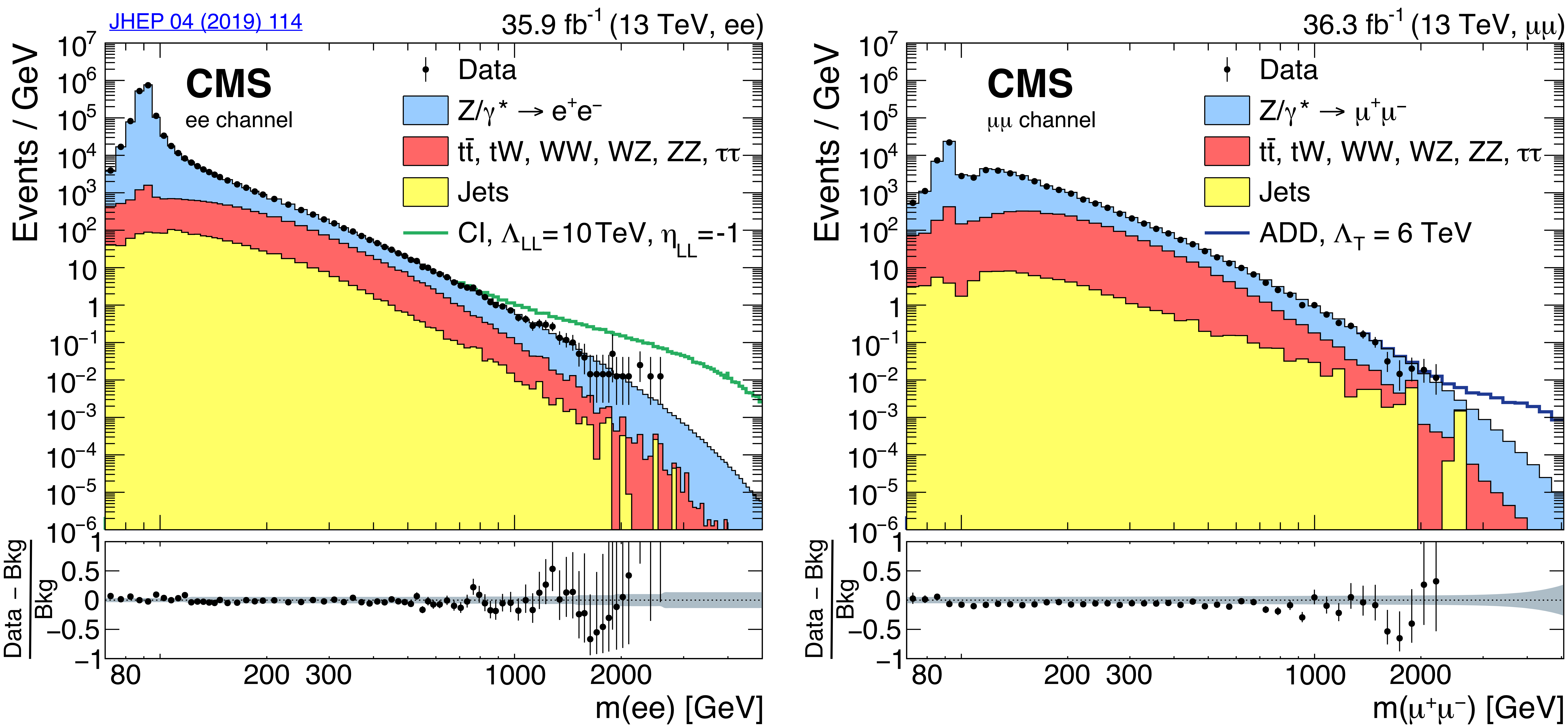
Source	Muon channel Background m_T = 2 (6) TeV
Trigger	1.1% (1.0%)
Lepton reconstruction and identification	8.9% (37%)
Lepton momentum scale and resolution	12% (47%)
E_T^{miss} resolution and scale	<0.5% (<0.5%)
Jet energy resolution	<0.5% (0.6%)
Multijet background	0.8% (1.5%)
Top-quark background	0.7% (<0.5%)
Diboson extrapolation	1.3% (9.7%)
PDF choice for DY	<0.5% (1.0%)
PDF variation for DY	7.4% (14%)
EW corrections for DY	3.7% (7.0%)
Luminosity	1.7% (1.7%)
Total	17% (62%)

$$x_{1,2} = \frac{M_{\ell\ell}}{\sqrt{s}} e^{\pm y}$$

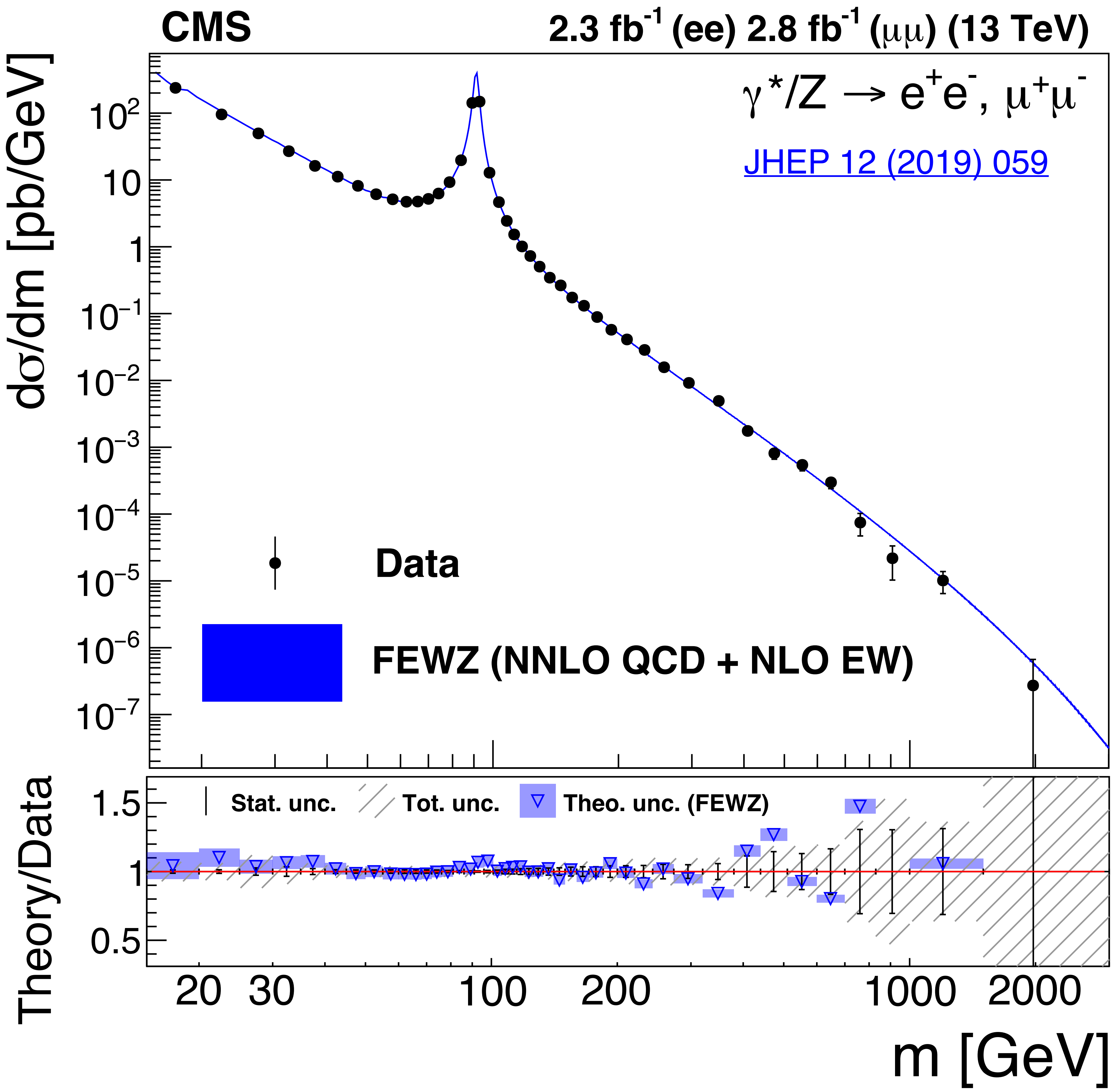


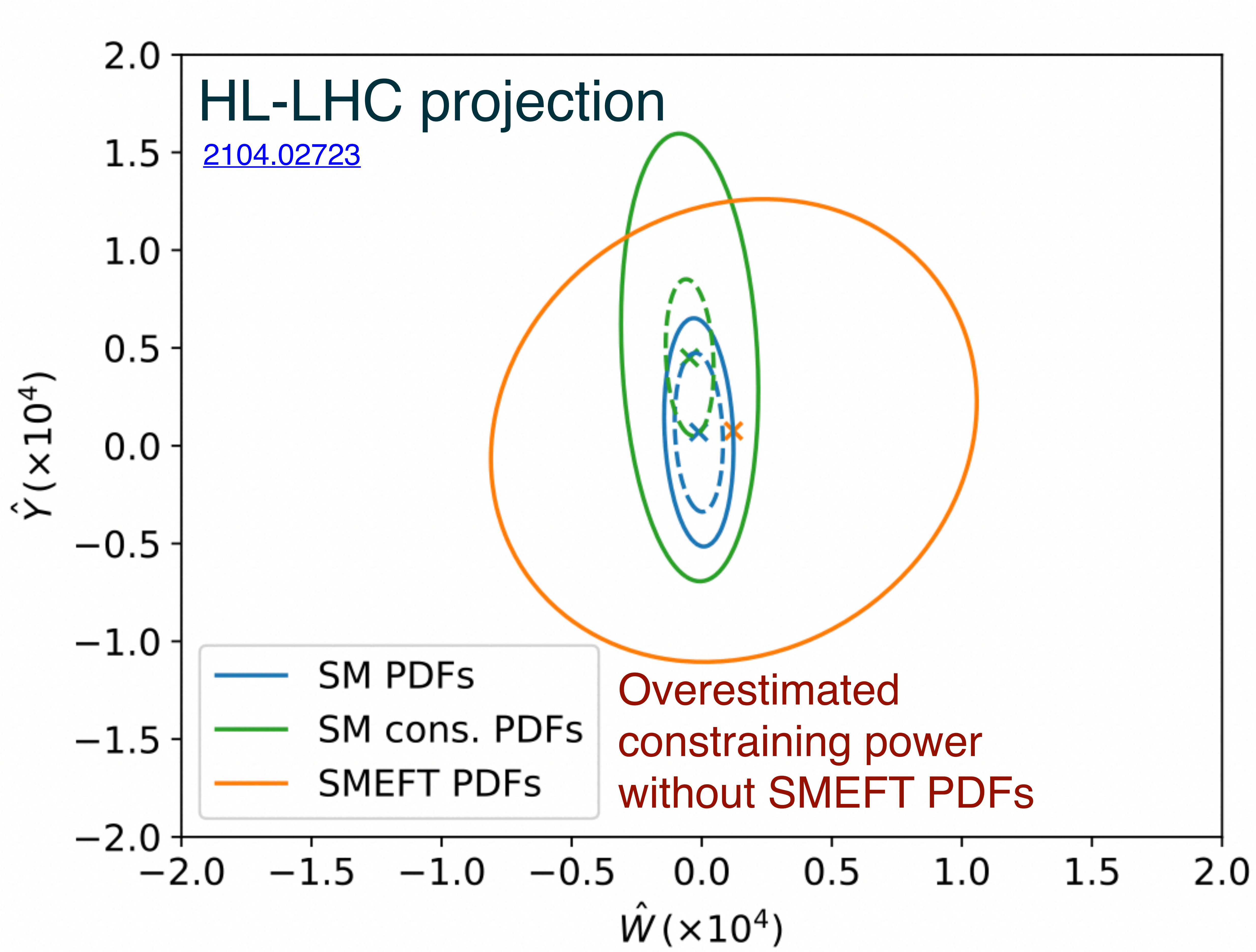
- Non-resonant signals at large energy scaled are even more susceptible to PDF uncertainties
- HMDY searches at LHC mostly rely on non-LHC data to constrain PDFs at high-x (e.g. HERA)
- However, for further progress, will also need constraints from the LHC high-x data
  - Need to be careful not to "tune" away potential new physics in LHC data with PDF fits!

JHEP 04 (2019) 114



- Searches can be optimized for specific signals, but they will be limited by the Drell—Yan theory uncertainties corresponding to the predictions that are available at the time
- On the other hand, unfolding the HMDY data allows for BSM interpretations in the future when the PDF uncertainties might be smaller— ATLAS / CMS measurements at ≥ 13 TeV available only with 2.8 fb^{-1}!
- **SMEFT PDF** (2104.02723): constraining PDFs and EFT parameters simultaneously will be essential for interpreting HL-LHC HMDY data in order to avoid "tuning away" new physics with PDFs





Summary

Summary

- Monte Carlo predictions for Standard Model backgrounds are essential for searches at the LHC
- NLO accurate predictions (e.g. for V+jets or ttbar) very often require combined data-MC procedures
- Likelihood fits with many normalization factors and nuisance parameters are performed to “tune” the Monte Carlo predictions in-situ to the desired phase space
  - Partially reduces the theory uncertainties, but only if the kinematics (shape) are well modeled
  - Increases statistical uncertainties (i.e. normalization factors)
- The most affected are searches for rare SM processes (e.g. rare Higgs decays, FCNC, …)
- Need to be very careful extracting EFT limits at high scale where PDF uncertainty becomes large

Backup