QCD background processes in BSM searches

Paolo Gunnellini
on behalf of the ATLAS and CMS Collaborations

August 2018 QCD@LHC 2018 Dresden (Germany)

Outline

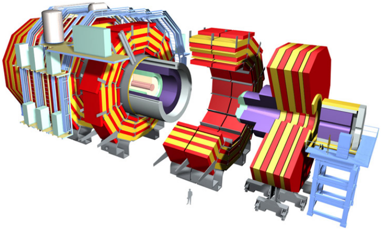
- General points on background estimation

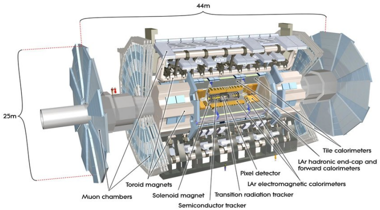
1. Bump hunt
2. ABCD, alpha and alphabet methods
3. Control regions and transfer functions
4. Template method
5. Parametrized extrapolation
6. Matrix element method (tight/loose ratio)

- Summary

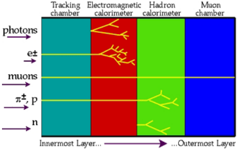
see talk by Pawel Klimek, Thursday 17.15h
see talk by Matthias Sampert, Thursday 17.40h

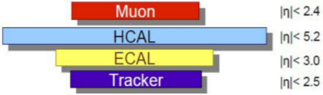
The CMS and ATLAS detectors





- Excellent performance on lepton identification
- Fine granularity and accurate single-particle reconstruction
- Very good jet energy resolution
- Hermetic coverage for reliable reconstruction of E_T^{miss}



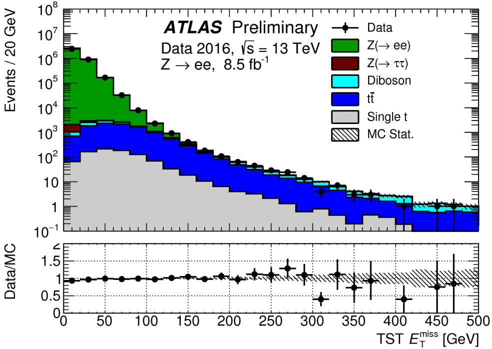


Understanding detector and predictions

→ Detector performance and measured observables generally well understood

→ Tremendous progress in Monte Carlo predictions

ATLAS-JETM-2016-008



Standard Model processes are backgrounds for searches and need to be carefully evaluated in the extreme regions of the phase space where we look for New Physics

→ Too large number of events to be simulated
→ Corners of phase space might be suboptimally predicted by available MC
→ Theory uncertainties become large when parton shower effects are relevant
DATA-DRIVEN BACKGROUND ESTIMATION METHODS

Which method to use?

Definition of control and validation regions

$\rightarrow$ Determination of background normalization
$\rightarrow$ Determination of background shape

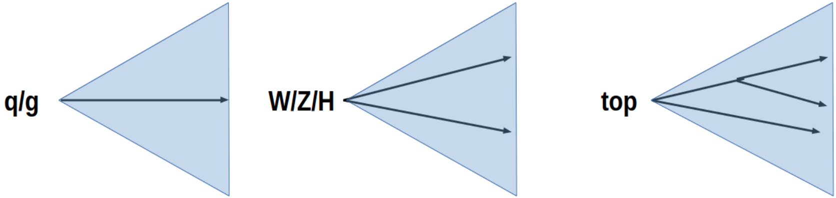
$\rightarrow$ Transfer function used for background evaluation in signal region

METHOD DEPENDS ON STUDIED SIGNAL

- **Resonance-like signal**
  1. Bump hunt methods
  2. Factorization cuts (ABCD / Alpha/ Alphabet methods)
- **Contribution on tail of distributions**
  1. Factorization cuts (ABCD / Alpha/ Alphabet methods)
  2. Parametrized extrapolation
  3. Template methods
  4. ... Matrix method, smearing (replacement) method

Selection is also crucial for increasing sensitivity to signal (trigger, jet substructure, specific taggers..)

Tagging boosted heavy objects against QCD



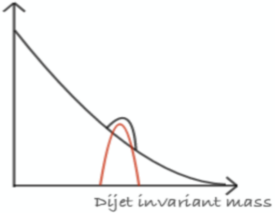
Large variety of techniques to discriminate the different substructure configurations

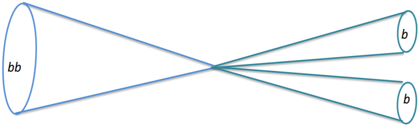
- large-cone jets (anti-k_T algorithm with $R = 0.8, 1.0$)
- removing soft radiation (pruning, grooming, soft-drop...)
- looking at the prongness of the fat jet (τ, ECF variables..)
- looking at the constituents ("deep" taggers)

BUT.. residual contribution from QCD is unavoidable

See talk Thursday morning by E. Ferreira De Lima and S. Marzani

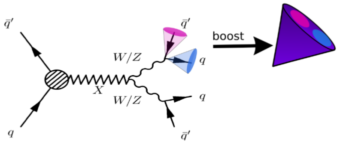
Methods for background estimation in searches for resonances decaying into heavy bosons or quark pairs

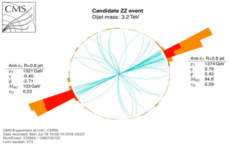




Bump hunt method

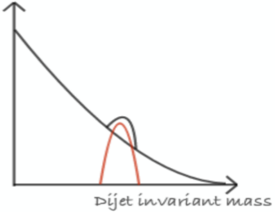
Fit measured distribution of a certain variable (e.g. dijet mass) and investigate the presence of a peak above the QCD background





Search for massive resonances decaying into VV or qV

One or two V-tagged (substructure-based) jets selected in the central region

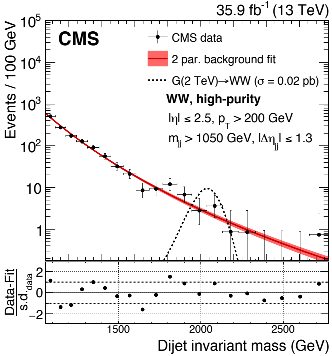


No control region used → Dijet invariant mass fit with background and signal shapes

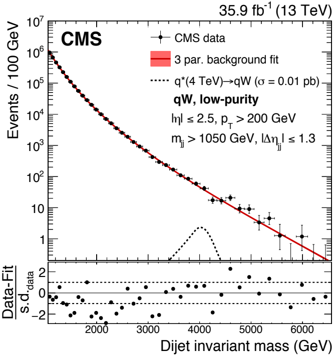
Uncertainties coming from different parametrizations of the background fit function

(CMS) PRD 97 (2018) 072006
(ATLAS) Phys. Lett. B 777 (2017) 91

Fit measured distribution of a certain variable (e.g. dijet mass) and investigate the presence of a peak above the QCD background



$$\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_1}}$$



$$\frac{dN}{dm_{jj}} = \frac{P_0(m_{jj}/\sqrt{s})^{P_2}}{(m_{jj}/\sqrt{s})^{P_1}}$$

Signal parametrized by a crystal ball function for various mass points

Fisher F-test drives the choice of the parametrization (CMS) PRD 97 (2018) 072006

Bump hunt method

Fit measured distribution of a certain variable (e.g. dijet mass) and investigate the presence of a peak above the QCD background

PROS:

- A completely data-driven method
- Easy to setup!
- Suitable for background falling spectrum and peaked signal
- It gives a clear and convincing proof of a discovery (presence of heavy resonance)

CONS:

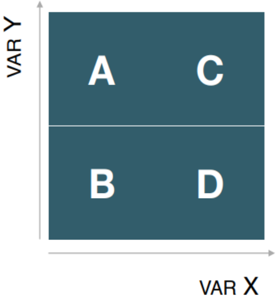
- Choice of fitted function and number of parameters is arbitrary
- Hard to parametrize turn-on in mass spectrum
- Not very robust at the end of the spectrum
- Very large width resonances may be absorbed in the fit
- Possible spurious signal contamination

Factorization cuts: ABCD method

By considering two uncorrelated variables, one splits a 2D phase space, in order to obtain a signal-like and a background-like region

If signal region is A:

$$N_A = N_B \frac{N_C}{N_D}$$



Assumptions:

- Signal contribution is negligible in regions B, C, D
- X variable has no impact on studied background
- X and Y should be uncorrelated

PROS: easy to setup + completely data-driven (which can be validated in MC)

CONS: difficult to test the lack of correlation between the two variables + needed extrapolation, which might be difficult to control

Bump hunt method tested on ABCD

$\sim$ 80 fb^{-1}, data recorded in 2015-2016-2017

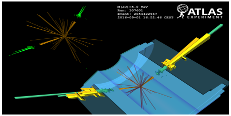
Two large-R (=1.0) jets with $\Delta y < 1.2$ and small p_T asymmetry

V-boson tagger based on jet mass and energy correlation function ratio D_2

Fit function used in signal region:

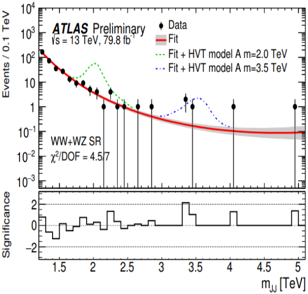
$$\frac{dN}{dm_{jj}} = p_1(1-x)^{p_2 - \epsilon p_3} x^{-p_3}$$

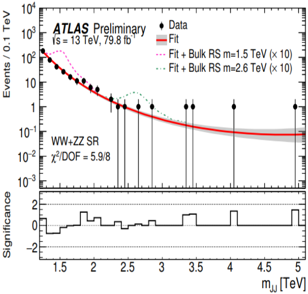
ABCD method applied as validation of the assumed parametric shape

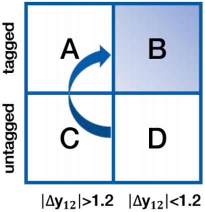


Search for diboson resonances decaying in hadronic final states

VAR X = Δy, VAR Y = V-tagging

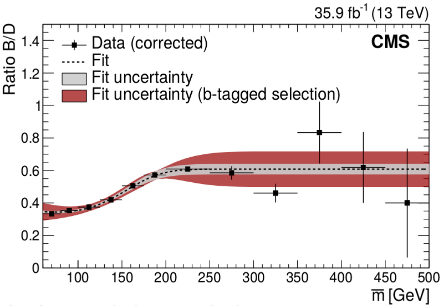


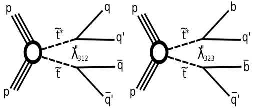




ATLAS-CONF-2018-016

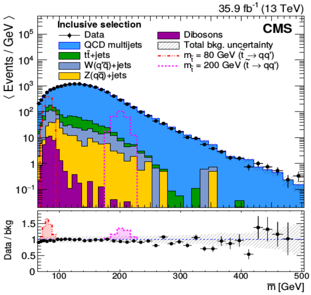
Factorization cuts: ABCD method

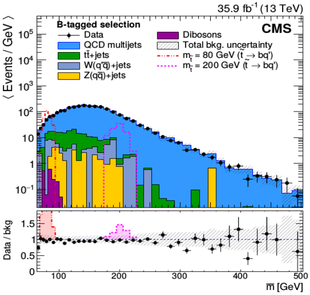


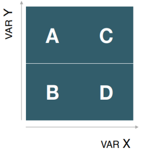


Search for pair-produced resonances decaying into quark pairs

VAR X $= m_{asym}$, VAR Y $= \Delta\eta$







	$m_{asym} < 0.1$	$m_{asym} > 0.1$
$\Delta\eta > 1.5$	Region *B*	Region *D*
$\Delta\eta < 1.5$	Region *A*	Region *C*

CMS-EXO-17-021
Subm. to PRD

Factorization cuts: alpha method

Definition of control region in order to estimate the background and evaluation of a transfer function (α ratio) to propagate these events to the signal region

Uncertainties driven by the statistical accuracy of the data in the control region

$$\alpha(x) = \frac{x_{MC}^{SR}}{x_{MC}^{CR}}$$

$$N_{bkg}^{SR}(x) = N_{data}^{CR} \cdot \alpha(x)$$

ASSUMPTIONS:

- Absence of signal in the control region
- MC simulation good in describing the measured x variable

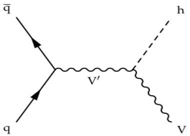
PROS: Easy, reliable and powerful (for an appropriate MC simulation)

CONS: Function estimated from MC might have large errors

EXAMPLE: Searches for heavy resonances decaying into VH pairs

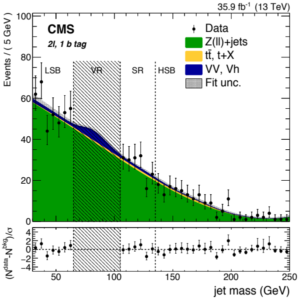
CMS-B2G-17-004 ATLAS-CERN-EP-2017-111

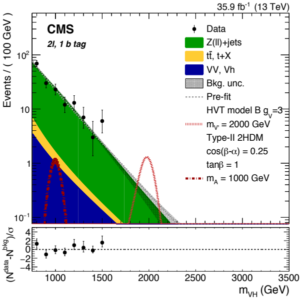
Definition of control region in order to estimate the background and evaluation of a transfer function (α ratio) to propagate these events to the signal region



Selection of two leptons from Z decay and one large-cone jet with 1 b-tagged subjet (Higgs candidate)

Main background consists of DY+jets processes





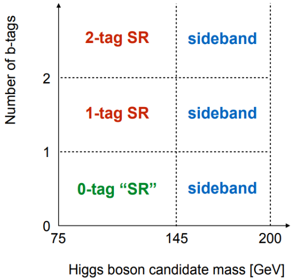
Alpha method for estimation of DY+jets background

$$\alpha(m_{VH}) = \frac{N_{SR}^{MC,bkg}(m_{VH})}{N_{SB}^{MC,bkg}(m_{VH})}$$

$$N_{SR}^{bkg} = [N_{SB}^{data} - N_{otherbkg}] \cdot \alpha$$

CMS-B2G-17-004
Subm. to JHEP

Factorization cuts: alpha method

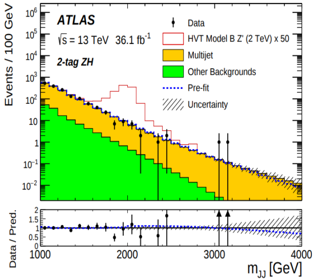


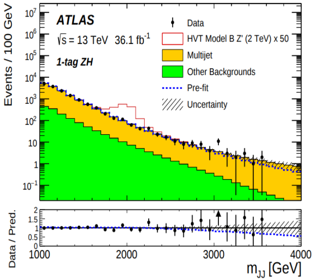
Search for heavy resonances decaying into VH ($\rightarrow$ qqbb)

Two large-R (1.0) jets at high p_T
(one V-tagged, one H-tagged)

Definition of a "0 b-tag" control region for multijet background estimation in 1-tag and 2-tag signal regions

$$\mu_{\text{Multijet}}^{1(2)\text{-tag}} = \frac{N_{\text{Multijet}}^{1(2)\text{-tag}}}{N_{\text{Multijet}}^{0\text{-tag}}} = \frac{N_{\text{data}}^{1(2)\text{-tag}} - N_{t\bar{t}}^{1(2)\text{-tag}} - N_{V+\text{jets}}^{1(2)\text{-tag}}}{N_{\text{data}}^{0\text{-tag}} - N_{t\bar{t}}^{0\text{-tag}} - N_{V+\text{jets}}^{0\text{-tag}}}$$



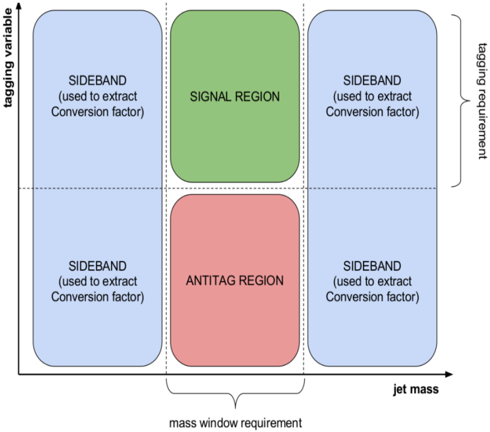


Background estimation improved by fits in the high-mass tail

PLB 774 (2017) 494

Factorization cuts: alphabet method

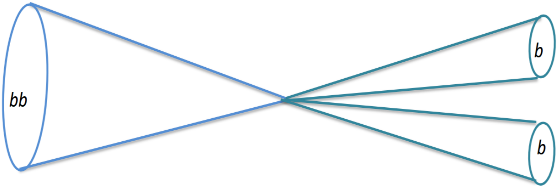
Use of multiple sidebands according to two variables, which encapsulate the signal region



Uncertainties coming from statistical uncertainty of the control regions

PROS:

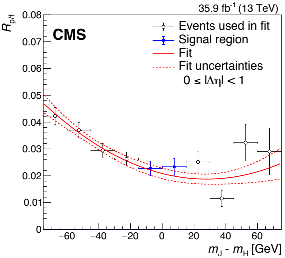
- It interpolates, instead of extrapolates (more robust)
- If enough statistics in control regions → reliable method also for regions with low statistics (tails)
- no assumption on shape

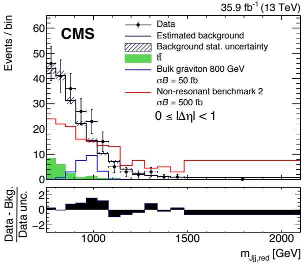


Search for Higgs boson pairs in four b-jet final states

Semi-resolved selection:
2 small-cone b-tag. jets
1 large-cone double b-tag. jet

Dominant background:
QCD multijet processes





Pass/fail ratio ($R_{p/f}$):

$$R_{p/f} = \frac{N_{pass}}{N_{fail}}$$

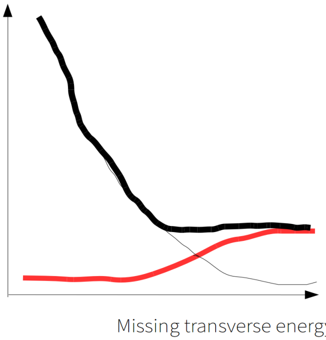
(on double b-tagger requirement)

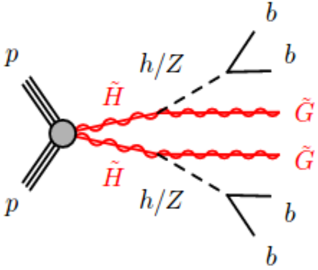
Estimated in sidebands (quadratic fit) and interpolated in the signal region

Background from $t\bar{t}$ taken from Monte Carlo simulation (@NLO)

CMS-B2G-17-019 - Subm. to JHEP

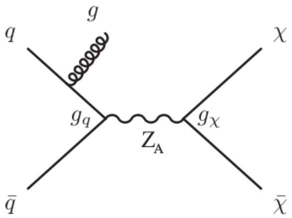
Methods for background estimation in searches for signals in tails of distributions





Normalization to control regions

Definition of control regions for estimation of background normalization



Search for new phenomena in events with one high-p_T jet and large missing transverse energy

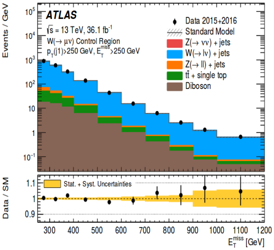
Main background processes are W+jets, Z($\rightarrow \nu\nu$)+jets, $t\bar{t}$ final states

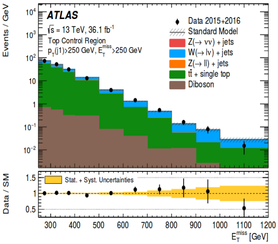
ATLAS - JHEP 01 (2018) 126
CMS - PRD 97 (2018) 092005

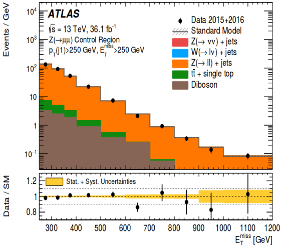
Definition of various control regions:
$\rightarrow$ One for each background process
$\rightarrow$ Using similar cuts as the signal region
$\rightarrow$ Simultaneous fit of background normalizations
$\rightarrow$ Shape taken from simulation

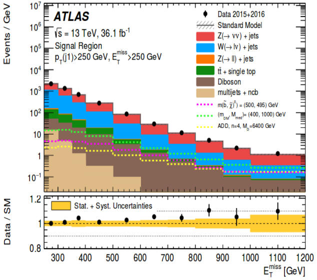
Correction factors: W/Z+jets(NLO): 1.27 $t\bar{t}$ (NLO): 1.31

Definition of control regions for estimation of background normalization







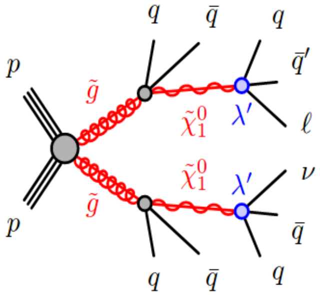


Simultaneous fit of E_T^{miss} in control regions

Extrapolation to signal region shows agreement with data

JHEP 01 (2018) 126

Factorization cuts: parametrized extrapolation



Search for new phenomena in multijet final states + lepton

→ Main background from W/Z+jets and $t\bar{t}$ processes

ATLAS (JHEP (2017) 88)
CMS (JHEP 05 (2018) 088)

Background normalization based on extrapolation from a lower jet multiplicity
→ "staircase-scaling" assumption
→ b-tag multiplicity distributions modelled by simulation

$$N_{j,b}^{W/Z+\text{jets}} = \frac{MC_{j,b}^{W/Z+\text{jets}}}{MC_{j}^{W/Z+\text{jets}}} \cdot N_5^{W/Z+\text{jets}} \cdot \prod_{j'=5}^{j'=j-1} r(j'),$$

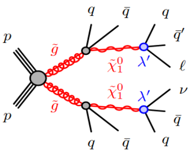
N_5 measured from data

$$r(j) = c_0 + c_1/(j+1),$$

c_0, c_1 measured from data

Uncertainties from statistical accuracy and modelling of b-tag multiplicity distribution

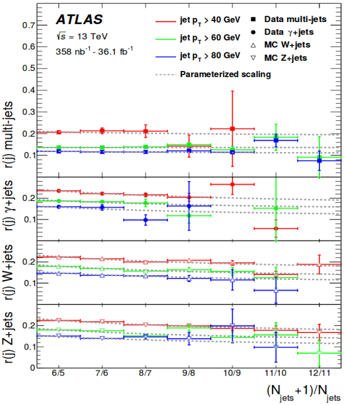
Factorization cuts: parametrized extrapolation

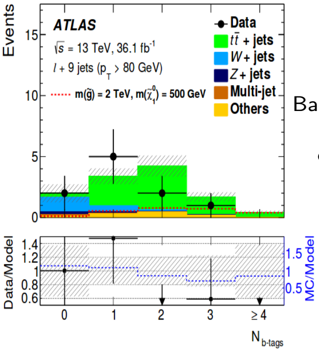


Search for new phenomena in multijet final states + lepton

ATLAS (JHEP (2017) 88)

Example: estimation of W/Z+jet background





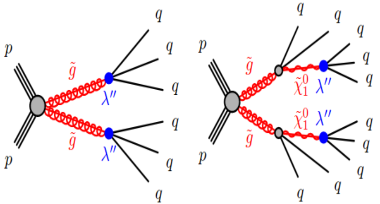
Stair-case scaling confirmed by data

Background extrapolation gives very good compatibility with data in signal region

Similar procedure for $t\bar{t}$ background estimation

Template method

→ Definition of a signal-depleted control region
→ Determination of the background shape in this region
→ Propagation of the shape in the signal region



Search for SUSY particles in multijet final states

→ Main background from QCD multijet processes

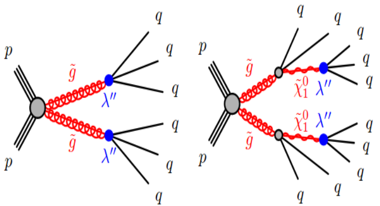
ATLAS (CERN-EP-2017-298) - Subm. to PLB

→ Selection based on observables from large-radius jets (R = 1.0)
→ Signal extraction from $M_J^{\Sigma} = \sum_{j=1..4} m_{jet}^j$

Single-jet templates as a function of jet p_T, η and b-quark content
→ assuming absence of correlation among jets

→ Uncertainties estimated through uncertainty determination regions (difference between predictions and observations)
→ Four validation regions and five overlapping signal regions

Template method



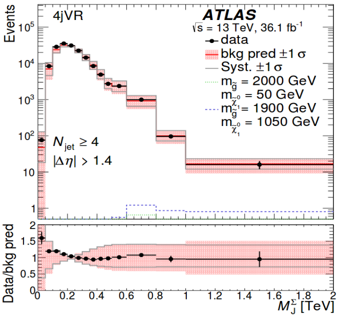
Search for SUSY particles in multijet final states

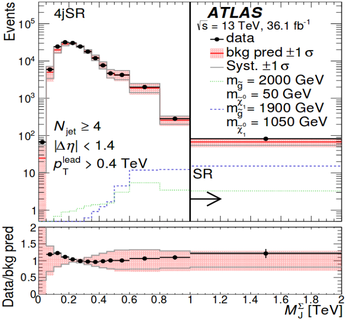
Control region requiring exactly three high-p_T jets

ATLAS (CERN-EP-2017-298) - Subm. to PLB

→ Selection based on observables from large-radius jets (R = 1.0)

→ Signal extraction from $M_J^{\Sigma} = \sum_{j=1..4} m_{jet}^{j}$





Templates are used for extracting M_J^{Σ} and normalized to the region between 0.2 and 0.6 TeV → looking for discrepancies at large M_J^{Σ}

Matrix method (tight/loose ratio)

Data-driven estimation of fake lepton identification from QCD multijet events

→ Definition of a lepton "tight-selection" and a "loose-selection" sample in QCD-enriched control regions

$$N^{loose} = N^{loose}_{real} + N^{loose}_{fake} \qquad N^{tight} = \epsilon_{real} \cdot N^{loose}_{real} + \epsilon_{fake} \cdot N^{loose}_{fake}$$

$$N^{tight}_{fake} = \frac{\epsilon_{fake}}{\epsilon_{real} - \epsilon_{fake}} (N^{loose} \epsilon_{real} - N^{tight})$$

Challenge: determination of ϵ_{real} and ϵ_{fake}

ϵ_{fake}: percentage of jets, selected with the loose lepton selection passing the tight selection

ϵ_{real} : percentage of loose leptons passing the tight selection

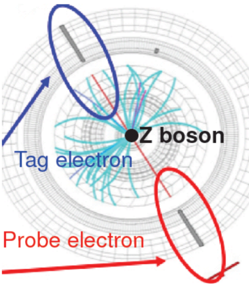
Search for gluinos with an isolated lepton, jets and E_T^{miss} - ATLAS - EPJC 76 (2016) 565

Matrix method (tight/loose ratio)

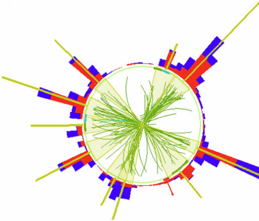
Determination of ϵ_{real}: tag-and-probe method

- Require a lepton-pair to be within the Z mass window
- Tag electron selected with tight selection
- Probe electron identified with loose selection

Measurement of the percentage of loose electrons which pass also the tight selection



Determination of ϵ_{fake}: QCD-dominated control regions with loose selection
→ measurement of percentage of jets that also pass the tight lepton selection
→ Selection in control region as independent as possible from the signal selection



PRO: can be used for a general number of final-state leptons
CON: possible overlaps between various backgrounds

Summary and overview

- LHC is delivering a large amount of data, which are being used for precision measurements and searches for new physics
- A large variety of clever and robust methods to estimate contributions from Standard Model processes directly from data is available and well understood
- The variety of methods allows cross check and their combination is useful to reduce systematic uncertainties/biases
- Input from theory more substantial for searches of non-resonant signals
- So far, no deviations from QCD/SM predictions in data even in remote corners of phase space

THANKS FOR YOUR ATTENTION

...and thanks to the EXO, SUSY, SMP, B2G conveners of the ATLAS and CMS collaborations for the help during the preparation of the slides!

QCD background processes in BSM searches: theory

Jonas M. Lindert

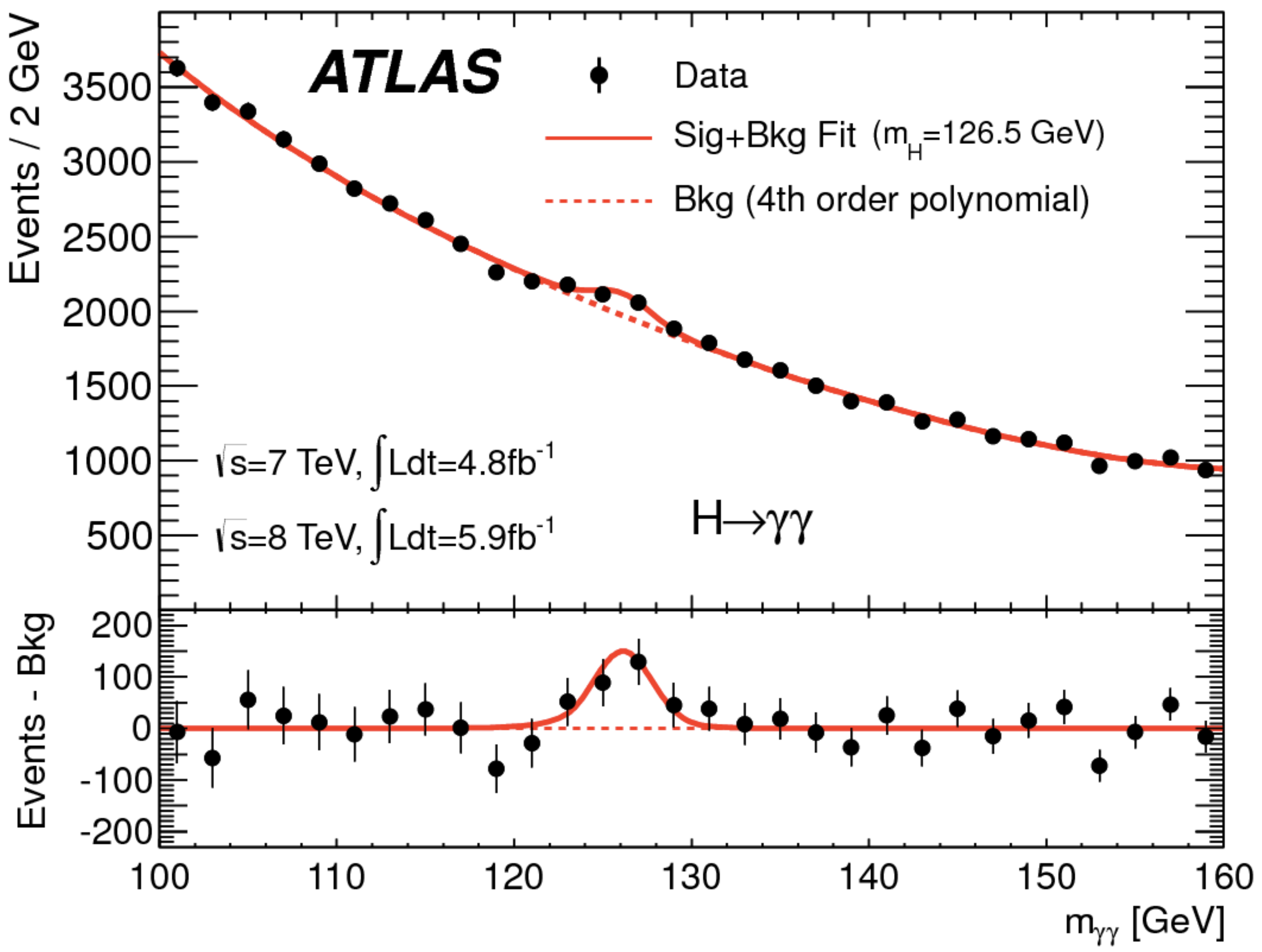
Durham University

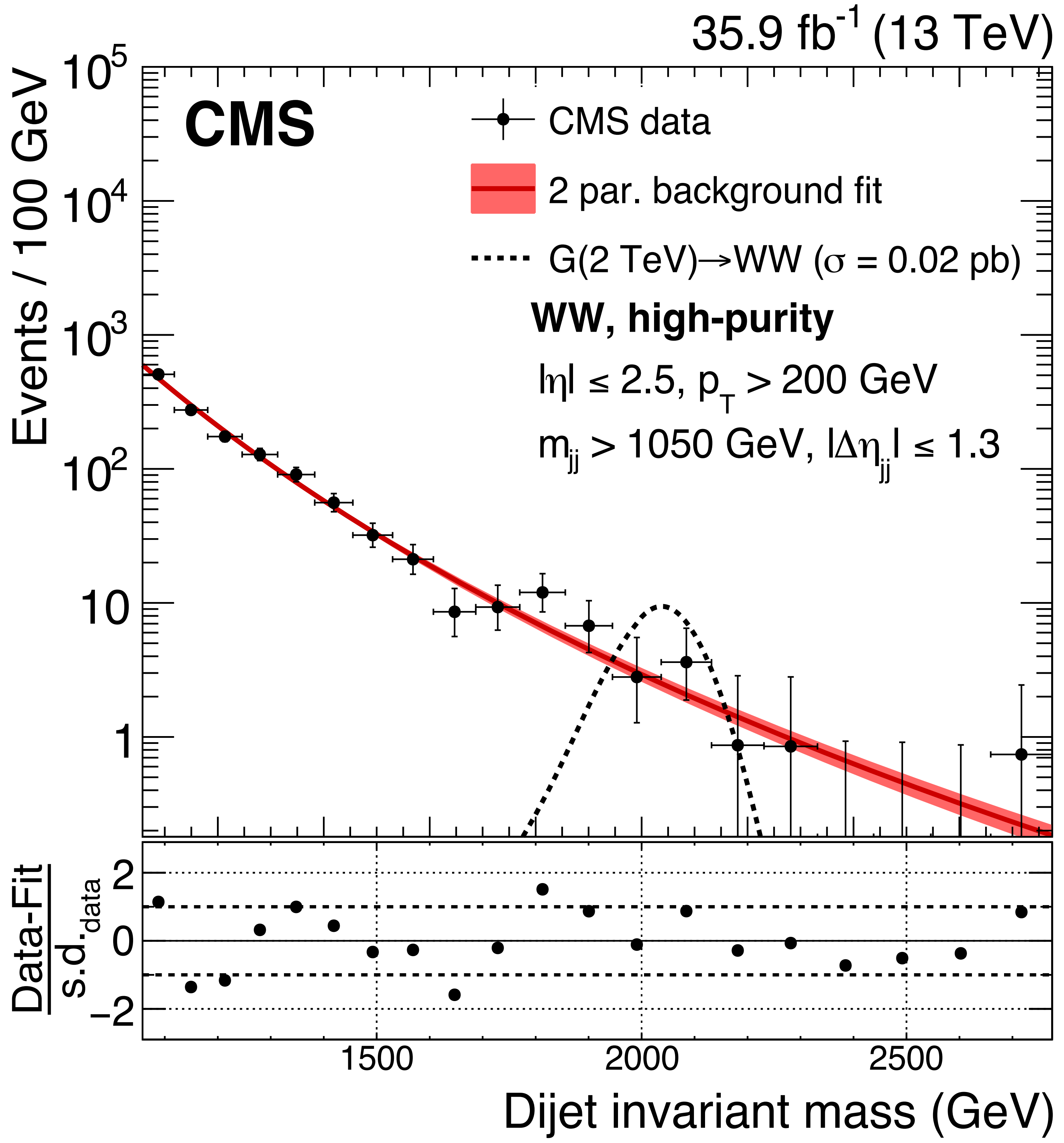
IP3

QCD@LHC

Dresden, 28. August 2018

From a theory/pheno perspective finding resonances is very easy…





$$\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_1}}$$

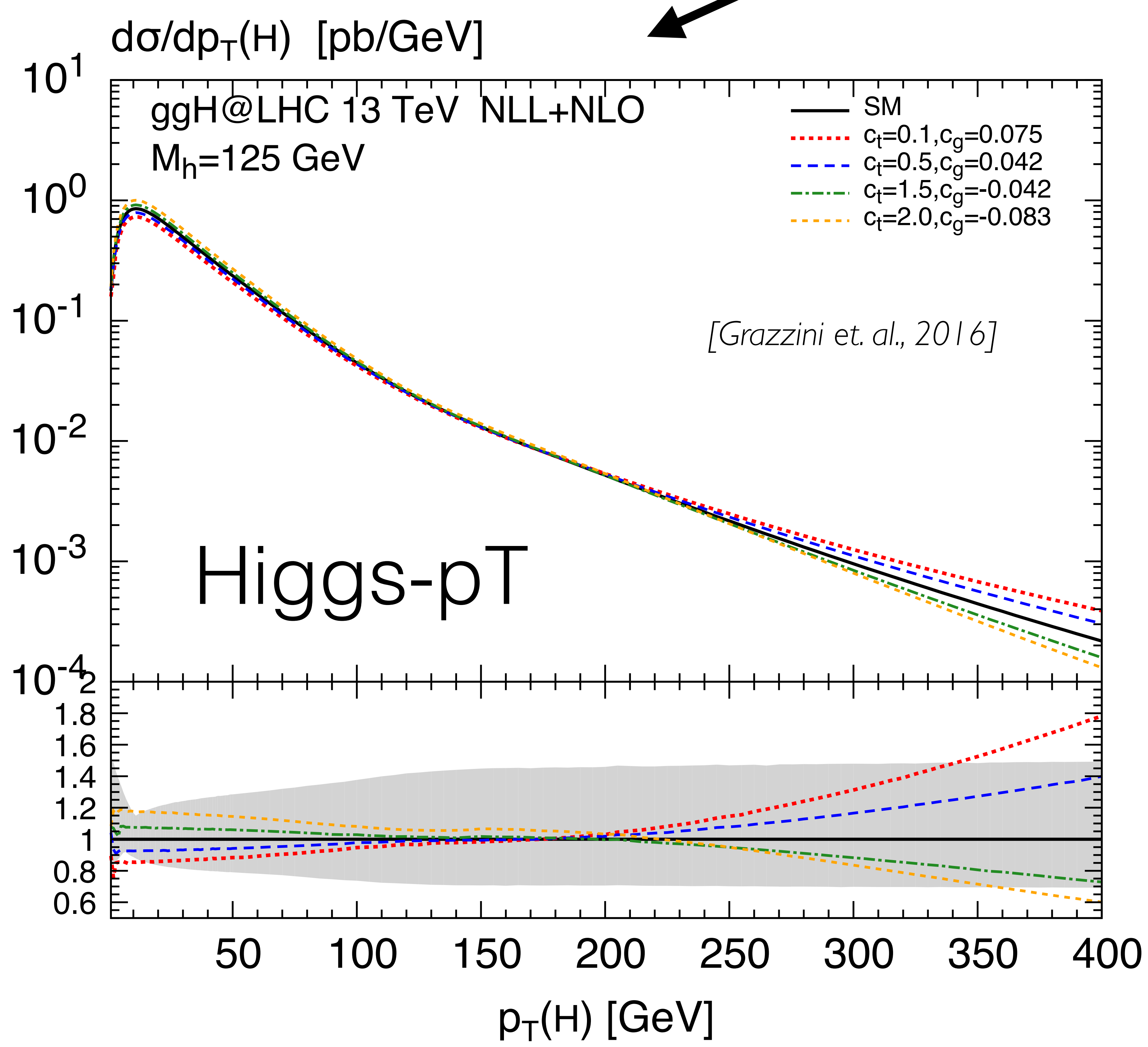
- Higgs at 125 GeV allowed for very clean discovery in **γγ** & 4l channels

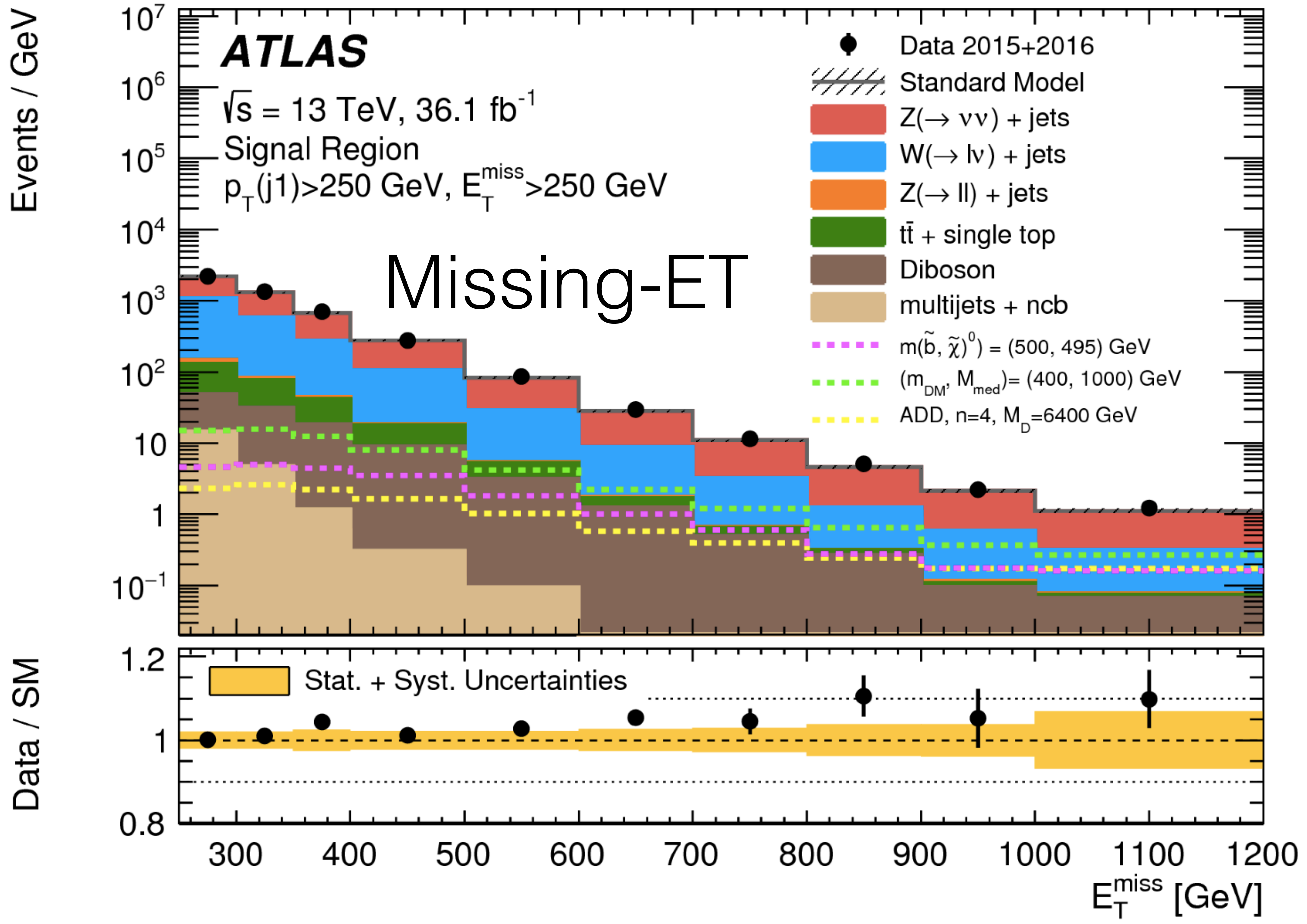
- Bump hunting: little to no theoretical input needed.

…finding new physics in tails of kinematic distributions in tough!

indirect searches ← Higgs

Dark Matter → direct searches





Look for BSM effects in small deviations from SM predictions:

→ Higgs processes natural place to look at

→ **good control on theory necessary!**

- Dark Matter particles produced at the LHC leave the detectors unobserved: signature missing transverse energy
- large irreducible SM backgrounds

→ **good control on theory necessary!**

Theoretical Predictions for the LHC

Hard (perturbative) scattering process

- **N(N)LO QCD + EW**

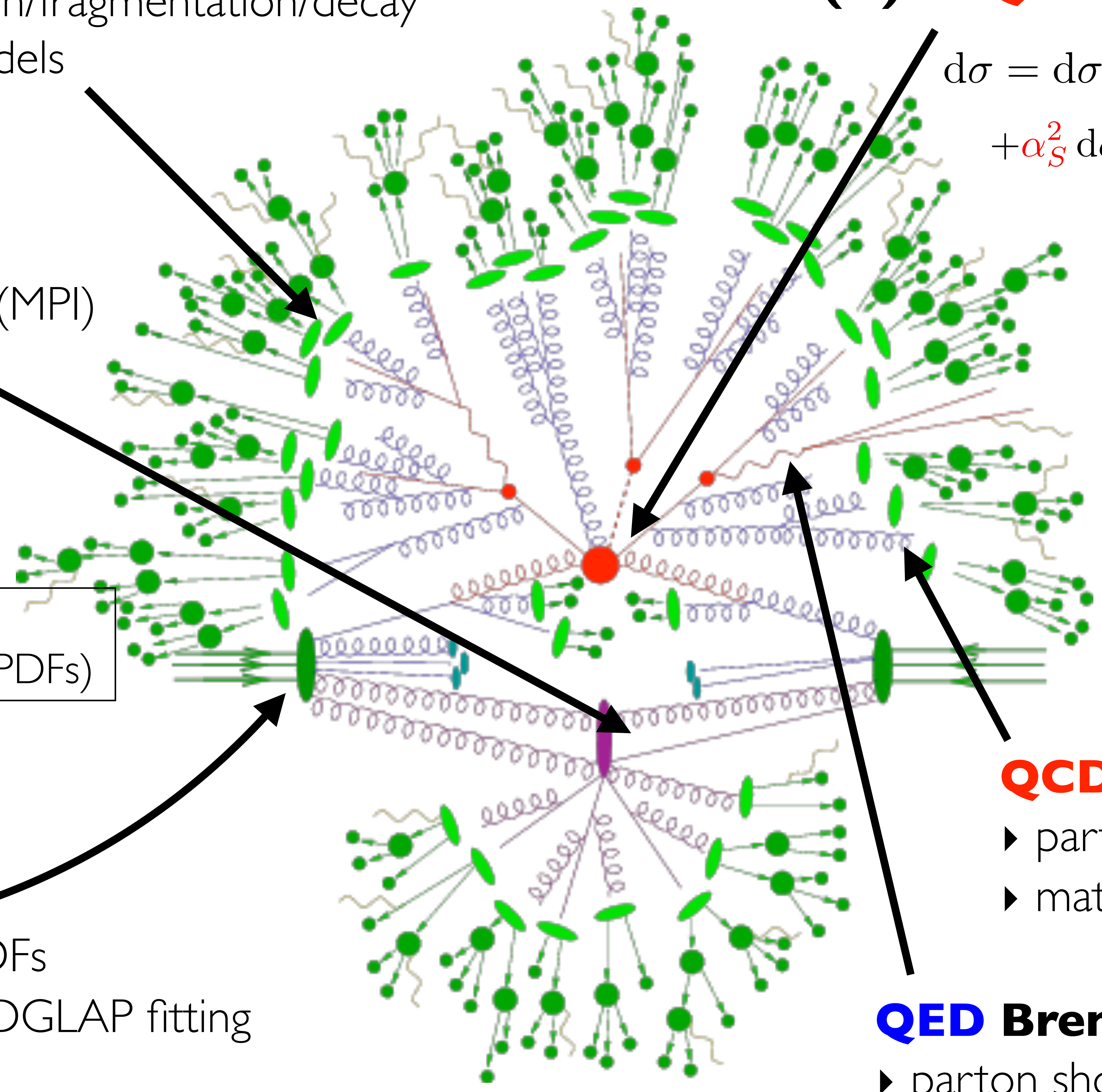
$$d\sigma = d\sigma_{\rm LO} + \alpha_S\, d\sigma_{\rm NLO} + \alpha_{\rm EW}\, d\sigma_{\rm NLO\,EW} + \alpha_S^2\, d\sigma_{\rm NNLO} + \alpha_{\rm EW}^2\, d\sigma_{\rm NNLO\,EW} + \alpha_S\alpha_{\rm EW}\, d\sigma_{\rm NNLO\,QCDxEW}$$

Hadronization/fragmentation/decay

- pheno models

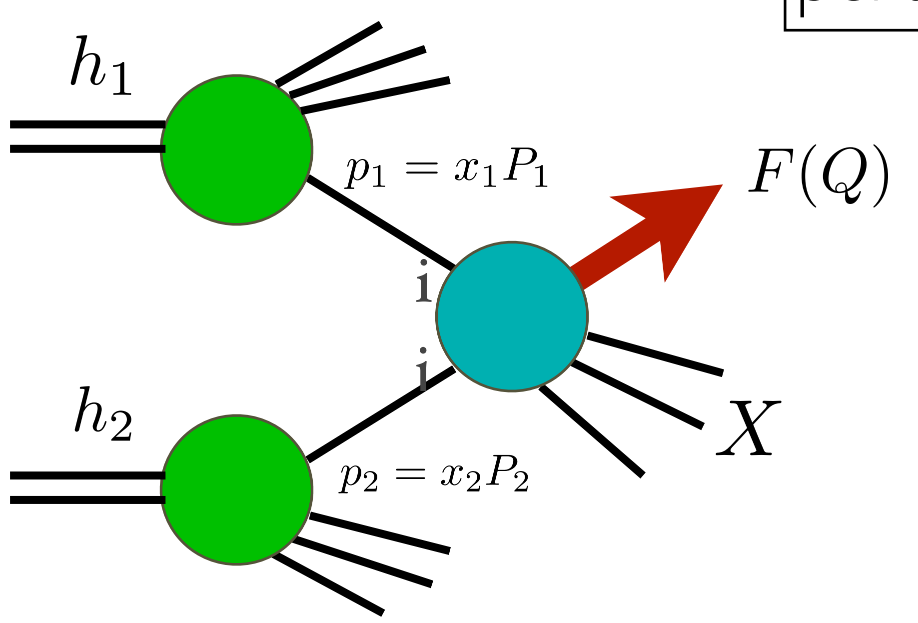
Multi Particle Interactions (MPI)

- pheno model



Key: QCD factorization:

Short distance non-perturbative effects (PDFs)



PDFs

- DGLAP fitting

QCD Bremsstrahlung

- parton shower
- matched to NLO matrix elements

QED Bremsstrahlung

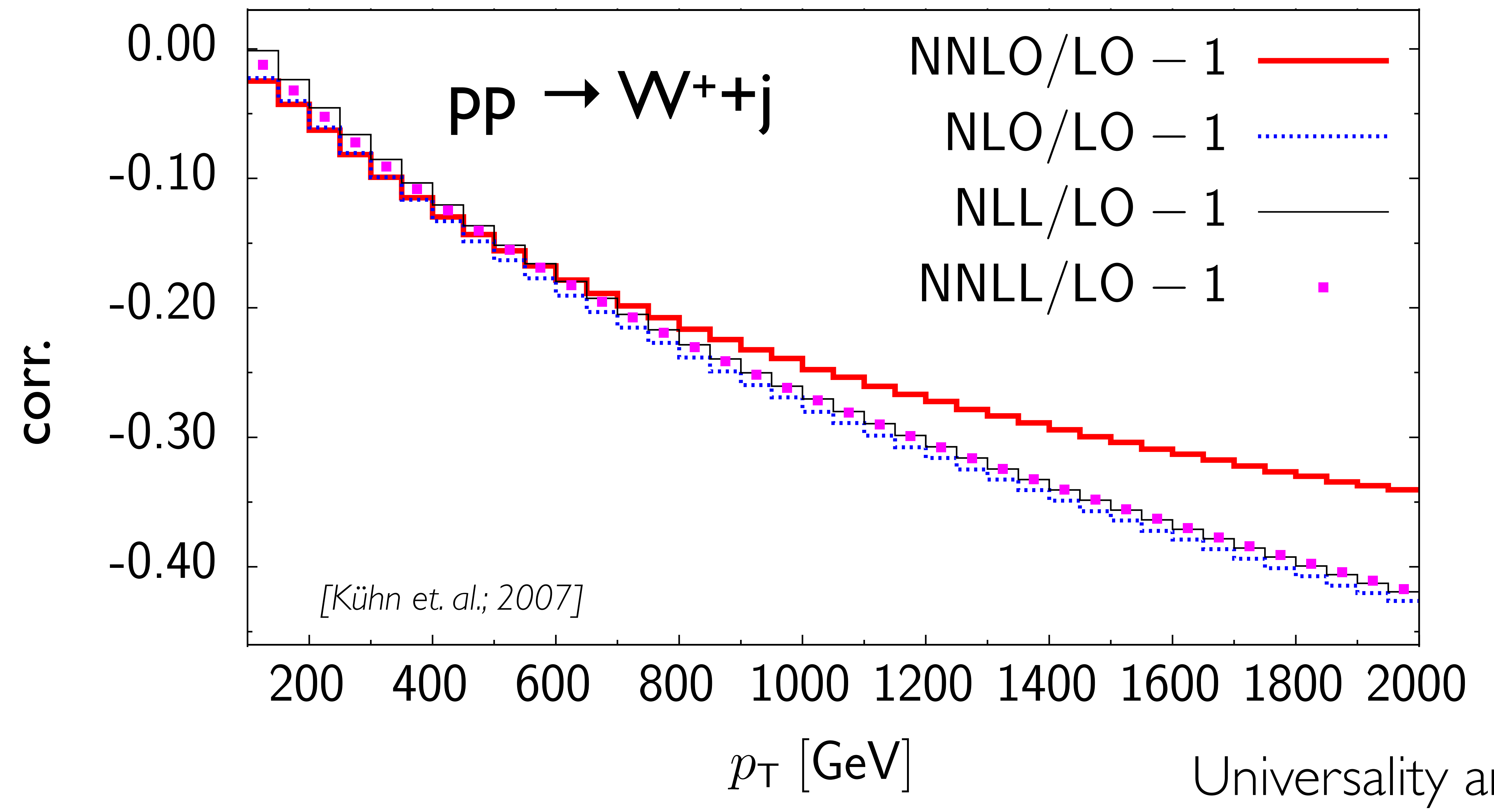
- parton shower
- matched to NLO matrix elements

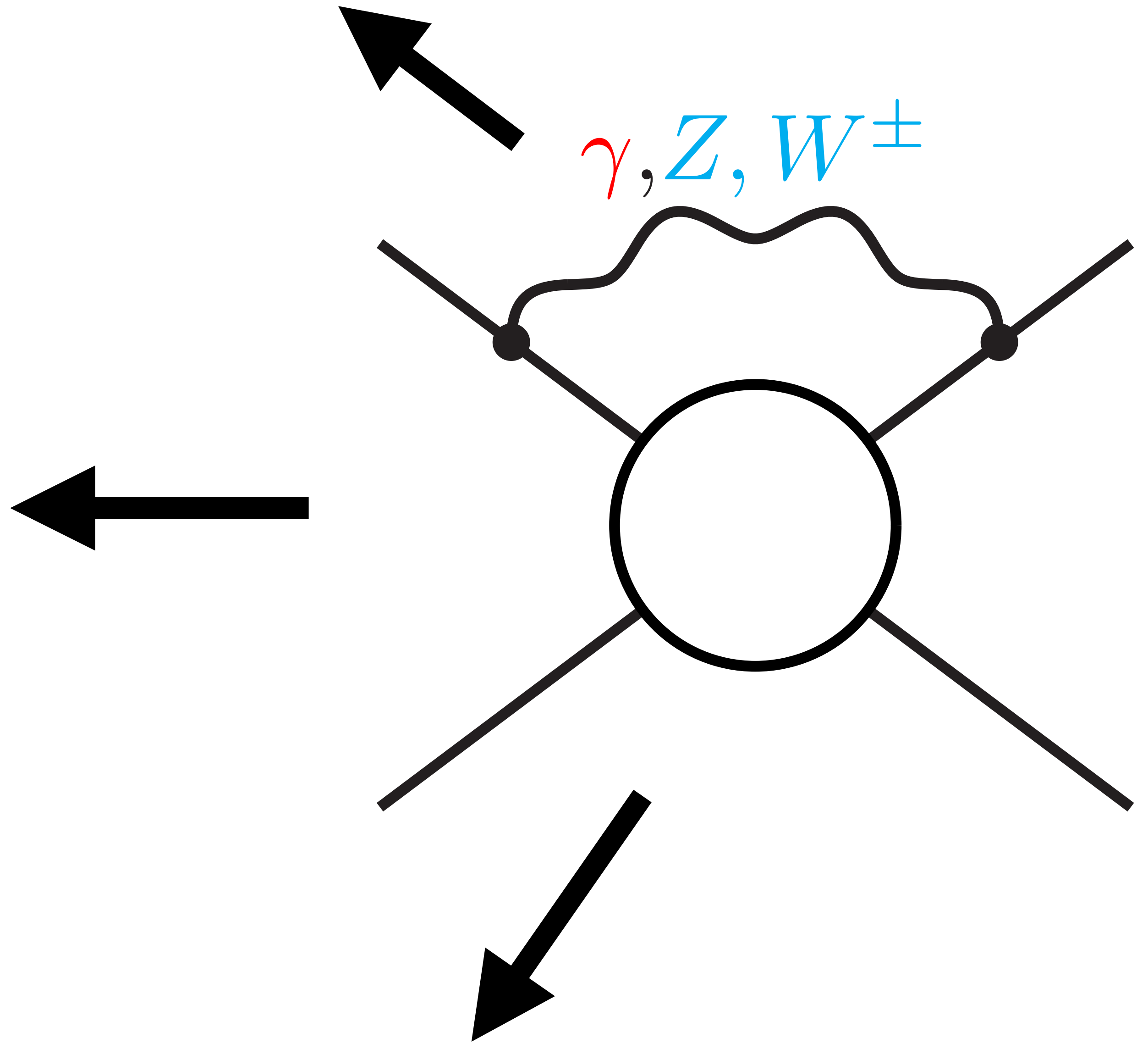
$$d\sigma = \sum_{ij} \int dx_1 dx_2 f_1^{(P_1)}(x_1) f_2^{(P_2)}(x_2) d\hat{\sigma}_{ij}(x_1 x_2 s)$$

Relevance of EW higher-order corrections in the tails of kinematic distributions

Numerically $\mathcal{O}(\alpha) \sim \mathcal{O}(\alpha_s^2) \Rightarrow$ **NLO EW ~ NNLO QCD**

Possible large (negative) enhancement due to soft/collinear **logs** from virtual EW gauge bosons:





[Ciafaloni, Comelli,'98; Lipatov, Fadin, Martin, Melles, '99; Kuehen, Penin, Smirnov, '99; Denner, Pozzorini, '00]

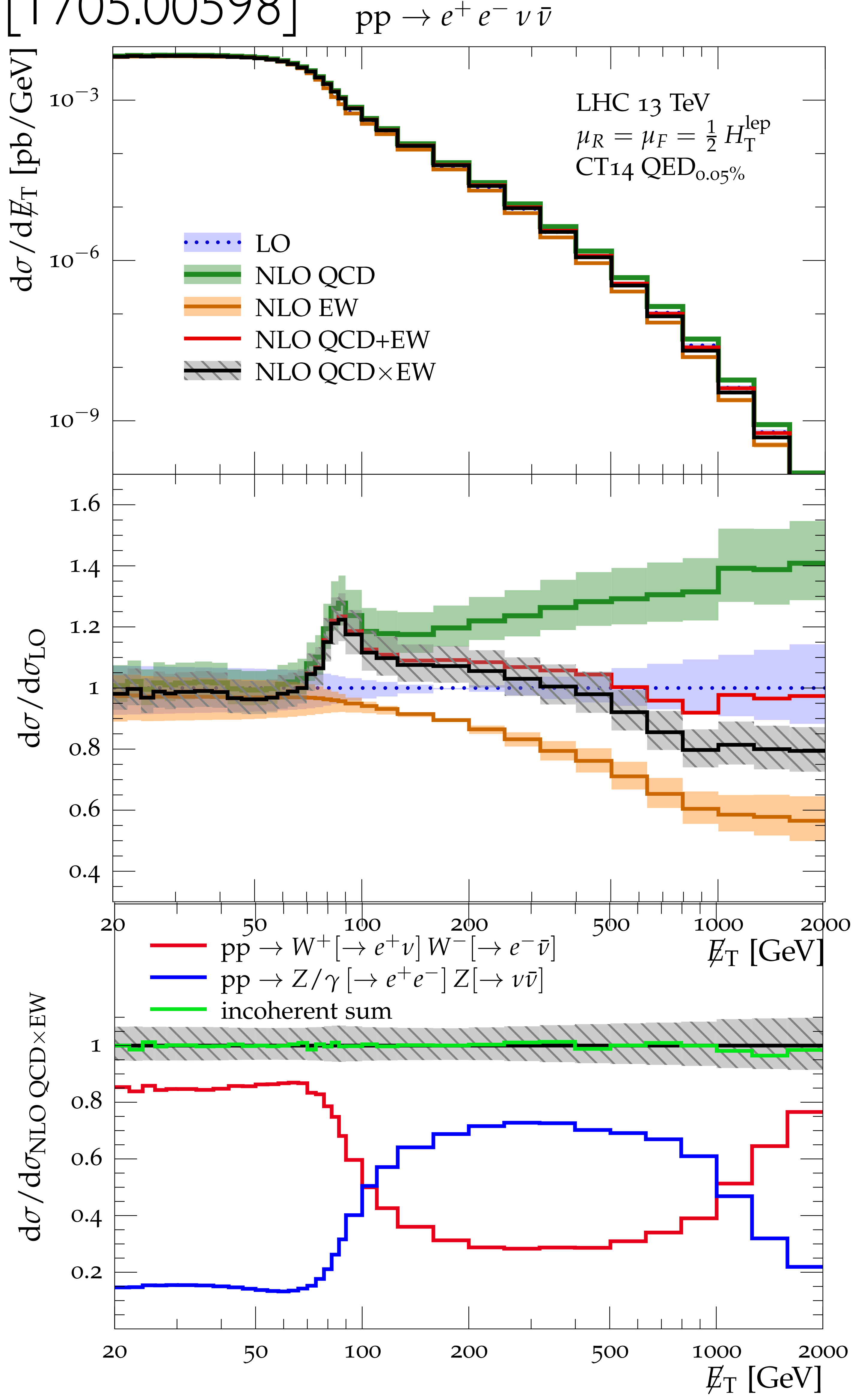
Universality and factorisation: *[Denner, Pozzorini; '01]*

$$\delta\mathcal{M}^{1-\mathrm{loop}}_{\mathrm{LL+NLL}} = \frac{\alpha}{4\pi}\sum_{k=1}^{n}\left\{\frac{1}{2}\sum_{l\neq k}\sum_{a=\gamma,Z,W^{\pm}} I^a(k) I^{\bar{a}}(l)\ln^2\frac{\hat{s}_{kl}}{M^2} + \gamma^{\mathrm{ew}}(k)\ln\frac{\hat{s}}{M^2}\right\}\mathcal{M}_0$$

➜ **overall large effect in the tails of distributions: p_T, m_{inv}, H_T,… (relevant for BSM searches!)**

off-shell vector-boson pair production at NLO QCD+EW

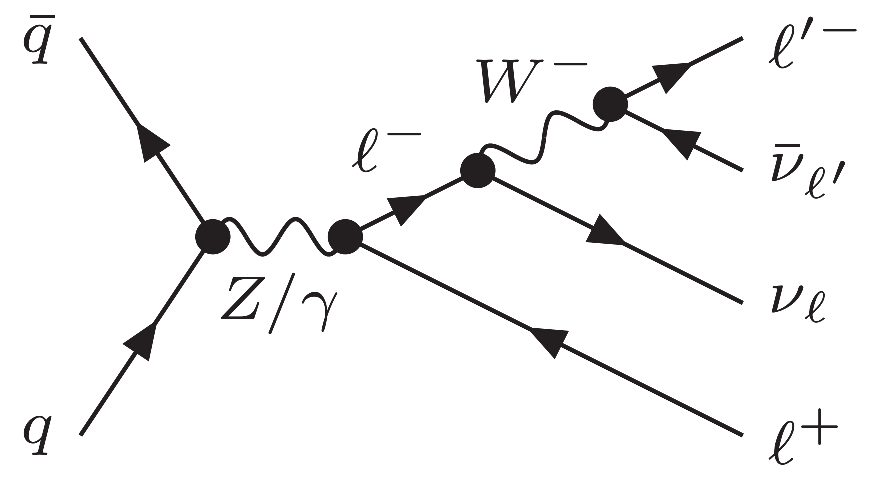
[1705.00598]



Effect of EW corrections strongly depends on the observable.

MET

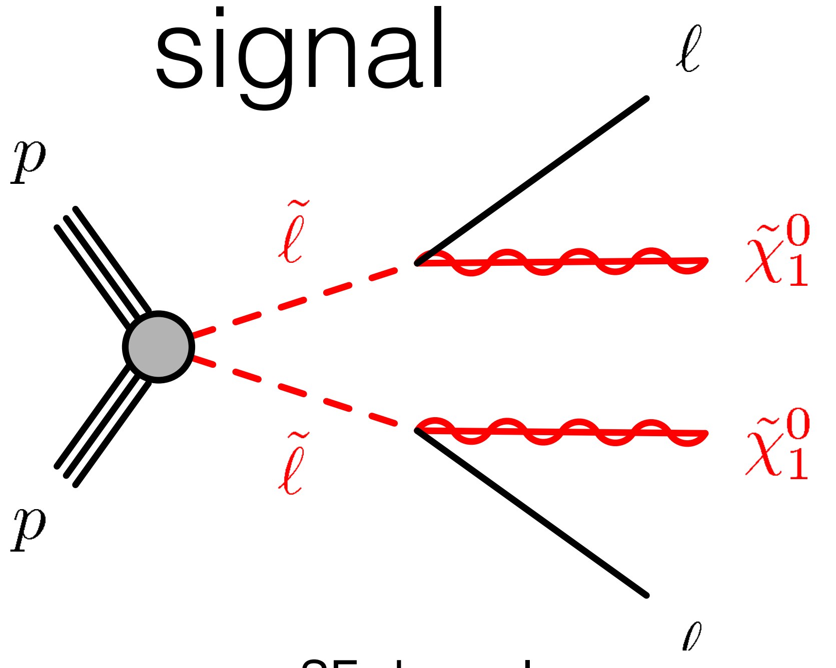
- at large MET>M_W: W's are forced off-shell



- jump in QCD corrections (extra jet unlocks back-to-back configuration)
- very large EW corrections: up to 50% (WW/ZZ dependent!)
- WW-ZZ interference very suppressed (as expected from LO)

off-shell vector-boson pair production at NLO QCD+EW

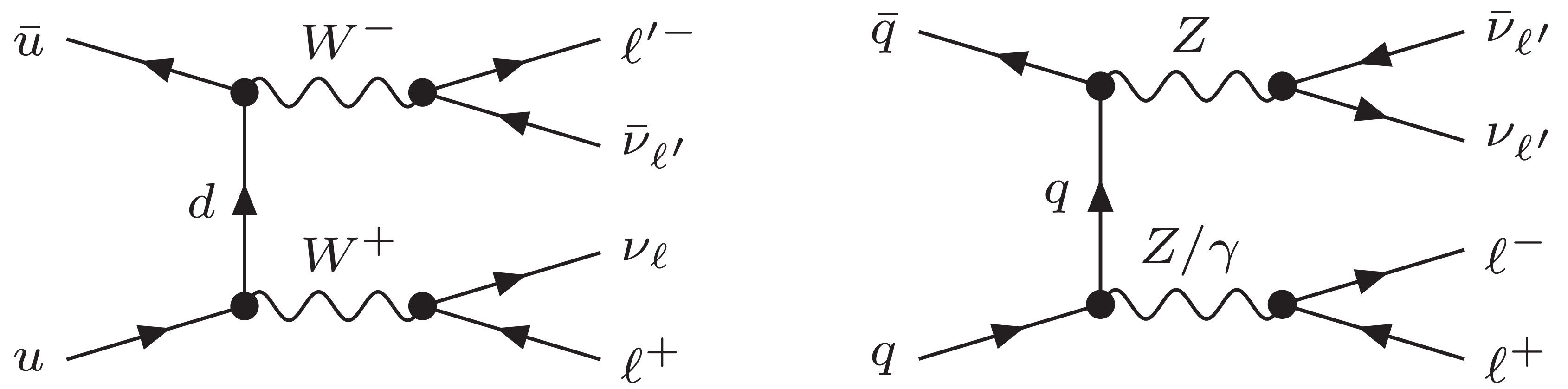
signal



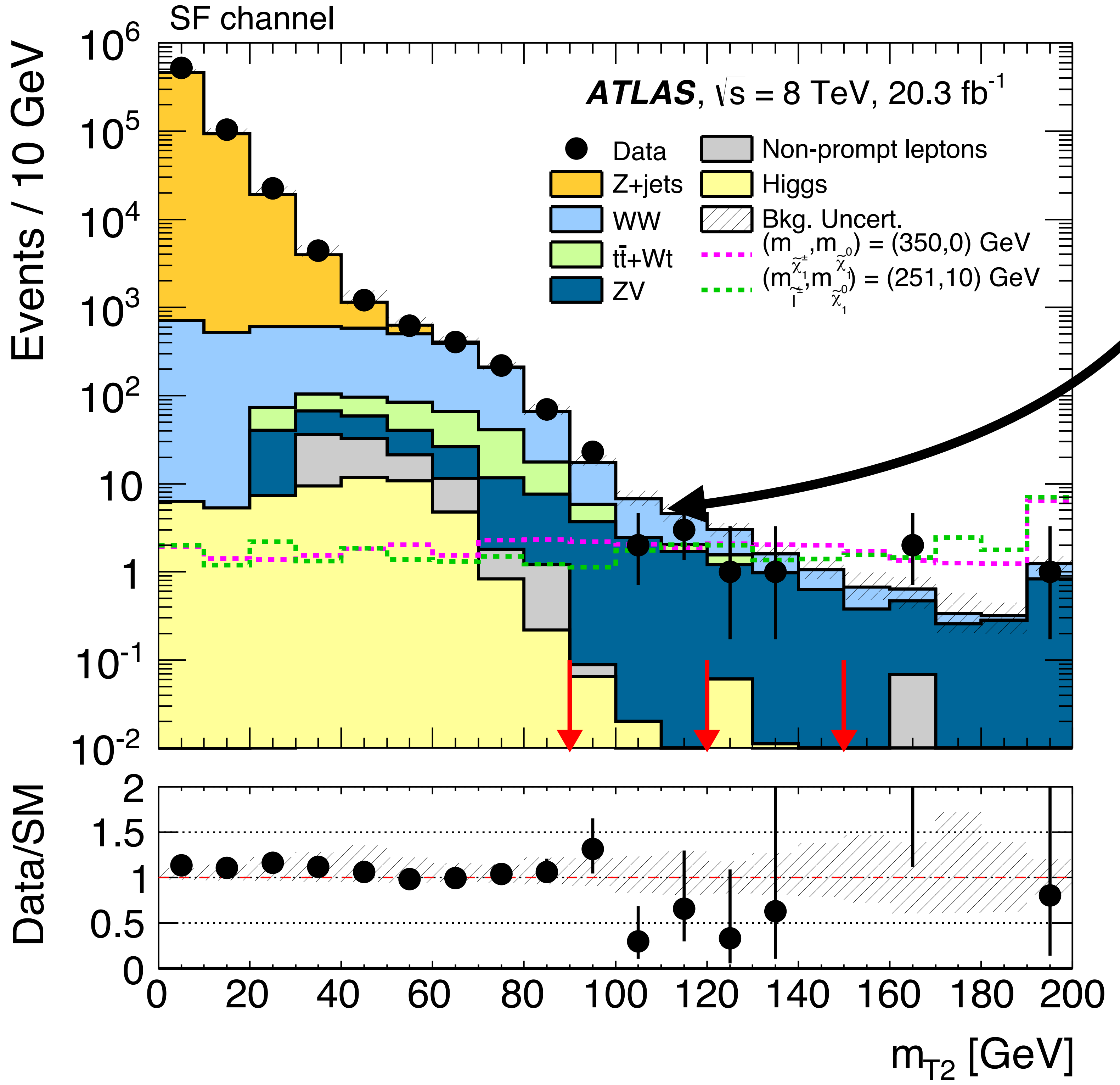
Direct Slepton pair production

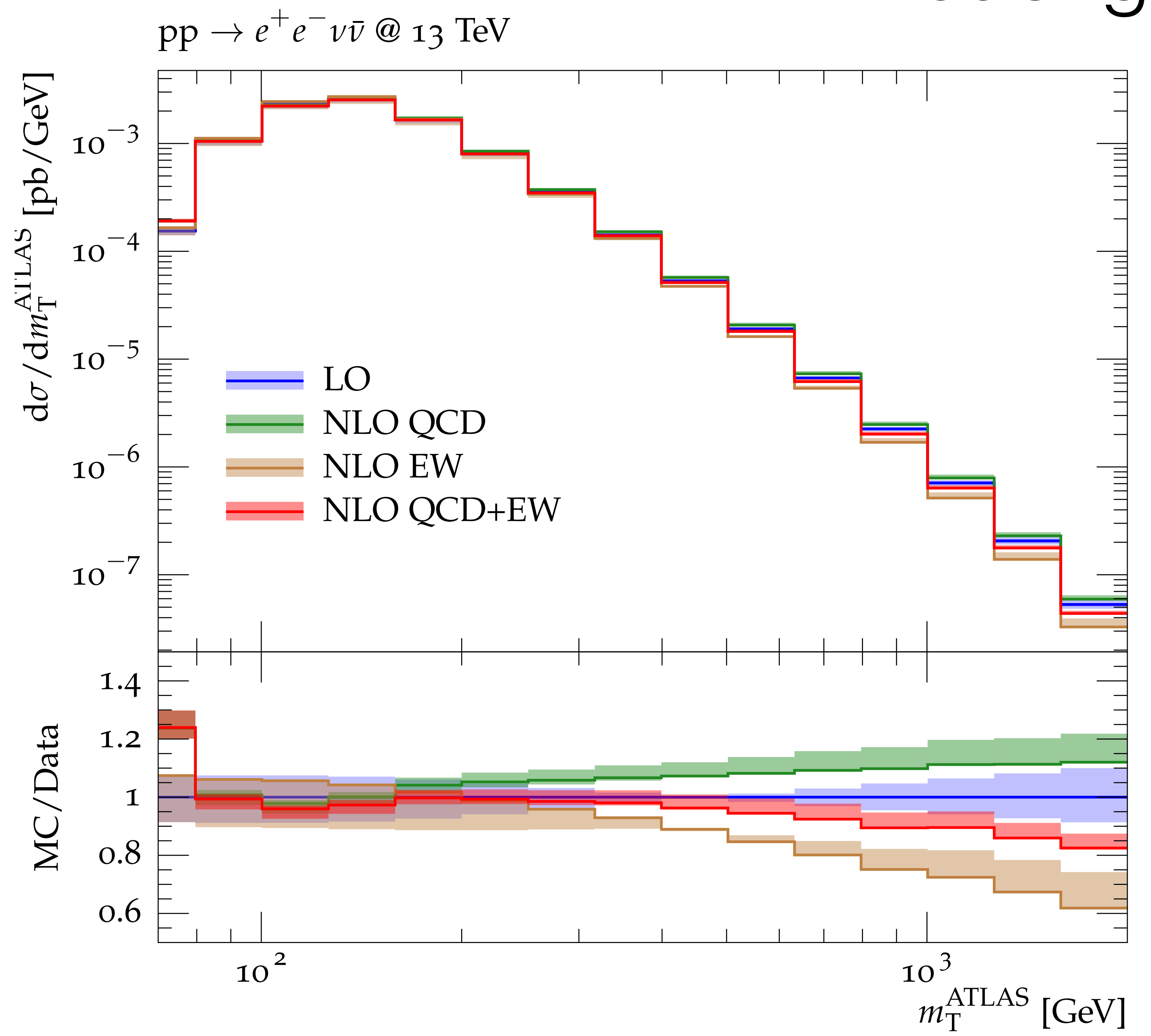
Signature: **2 OS-SF leptons + MET**

Background: $W^+W^-/ZZ \rightarrow 2l2\nu$



background



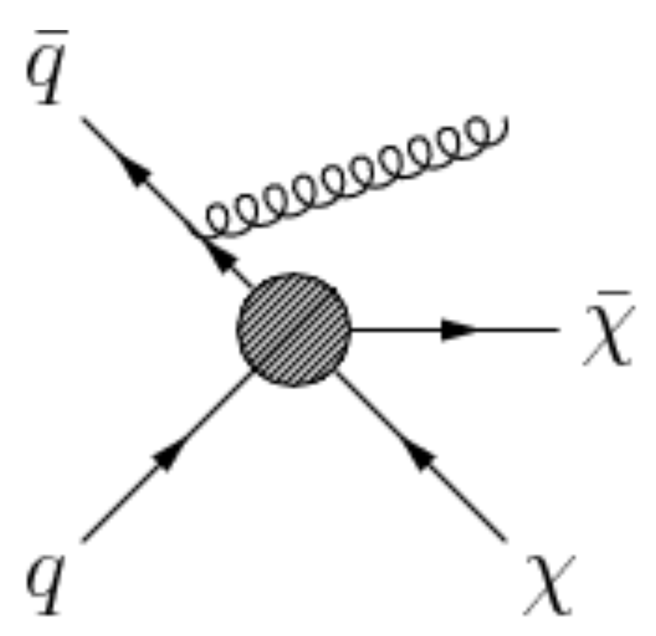


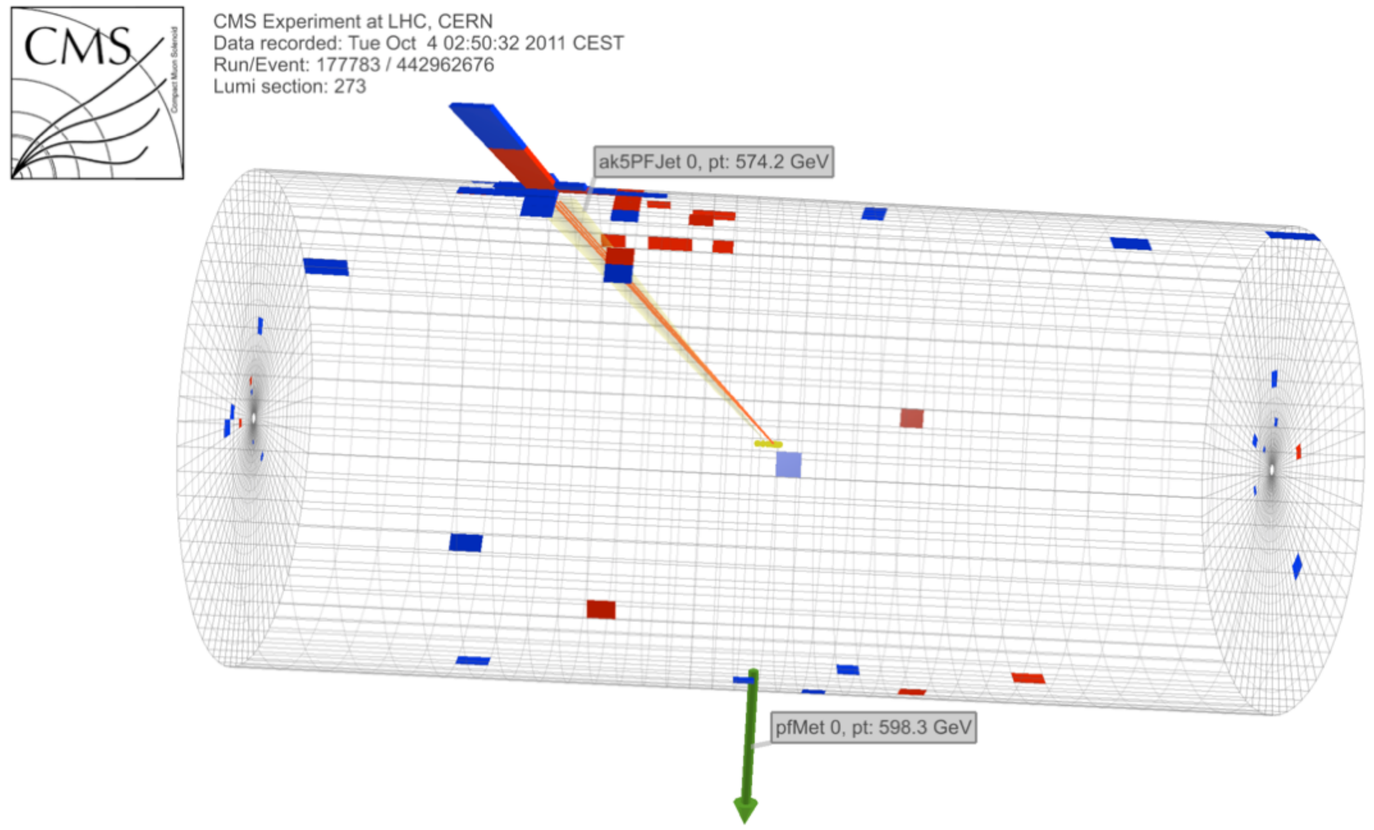
- VV background entirely from MC
- validated in control regions

- EW corrections: -10(-20)% for mT2=200-300 GeV(1 TeV)
- important to include in the future to avoid fake signals!

SM backgrounds in monojet / MET+X searches

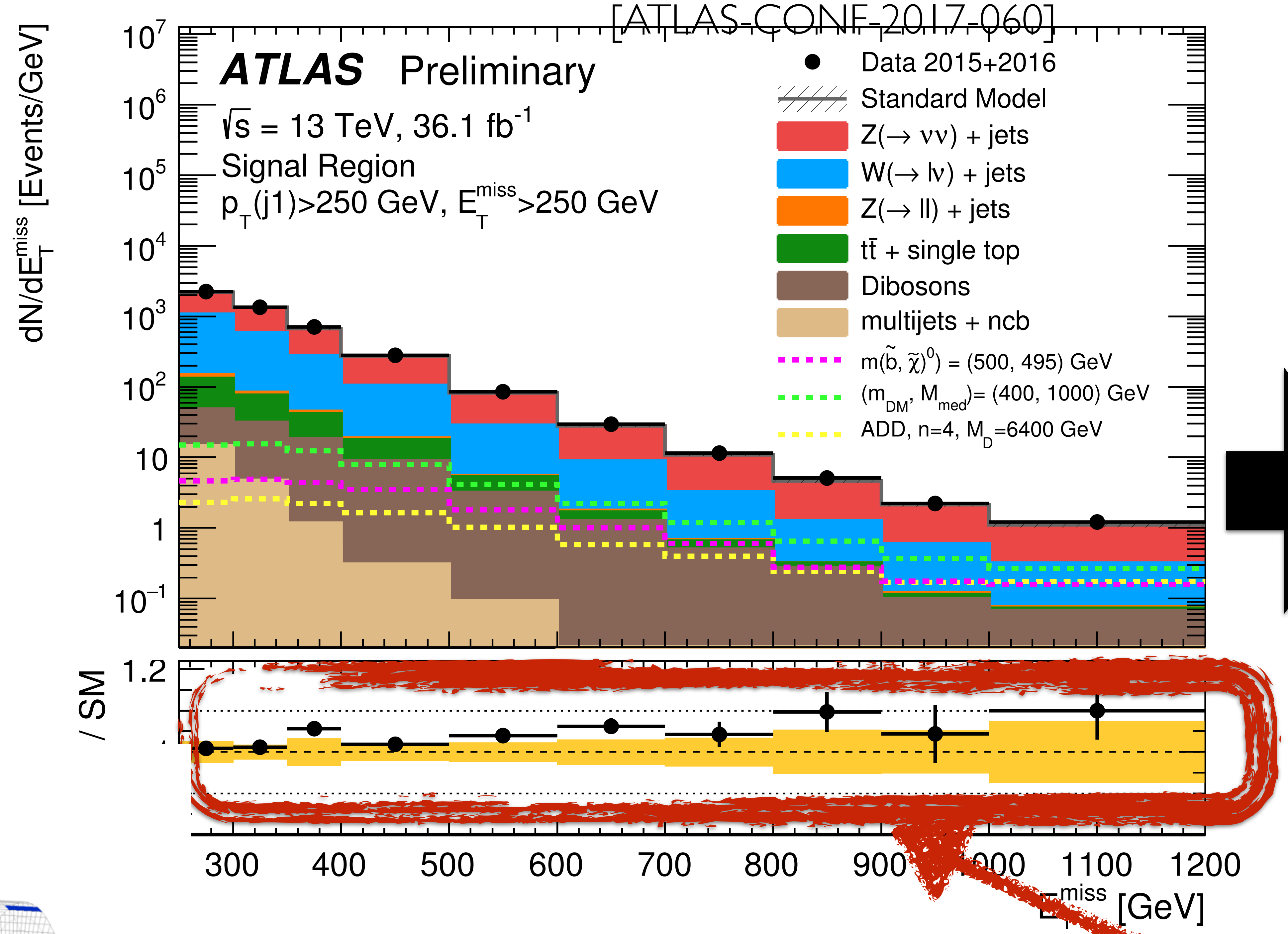
we hope to see:

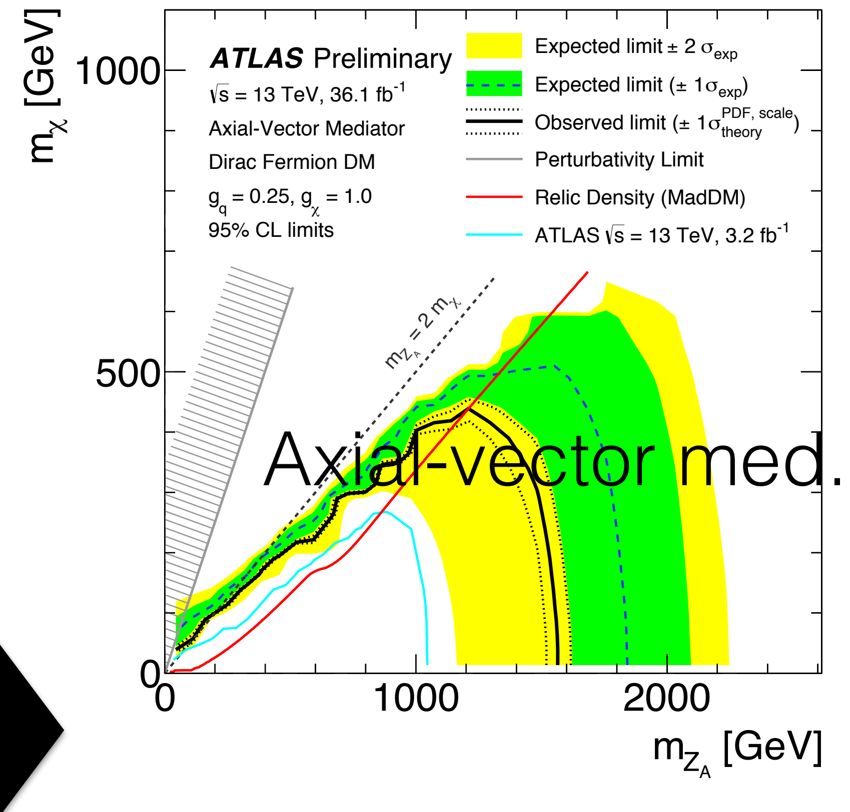




missing transverse energy

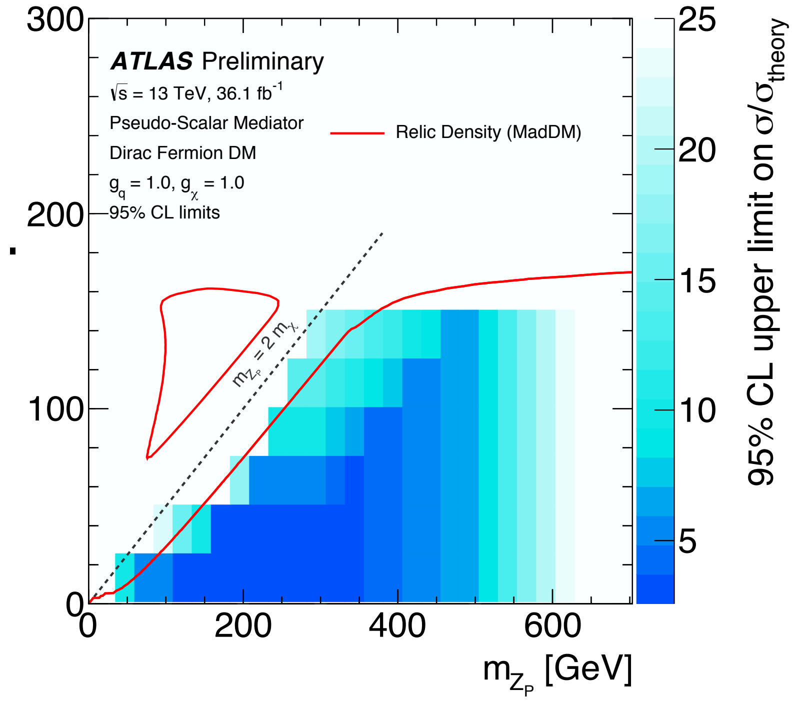
[ATLAS-CONF-2017-060]

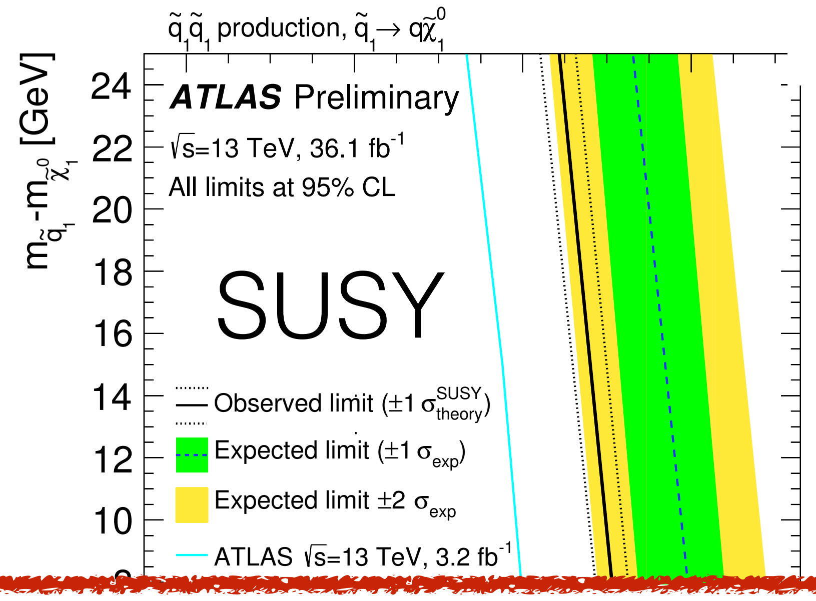




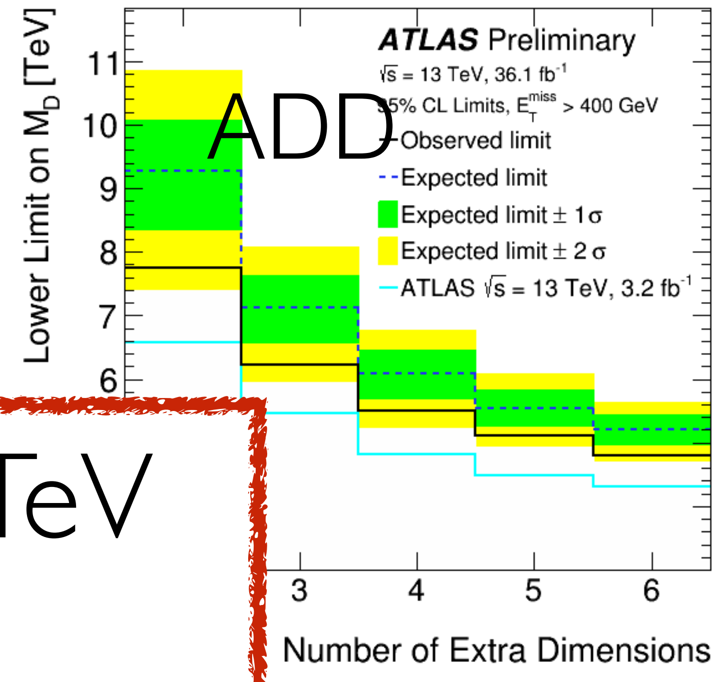
Axial-vector med.

Pseudoscalar Med.





SUSY

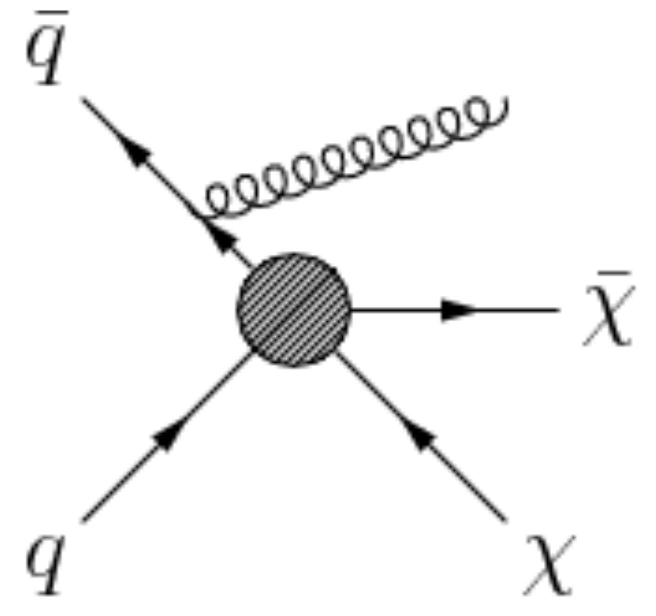


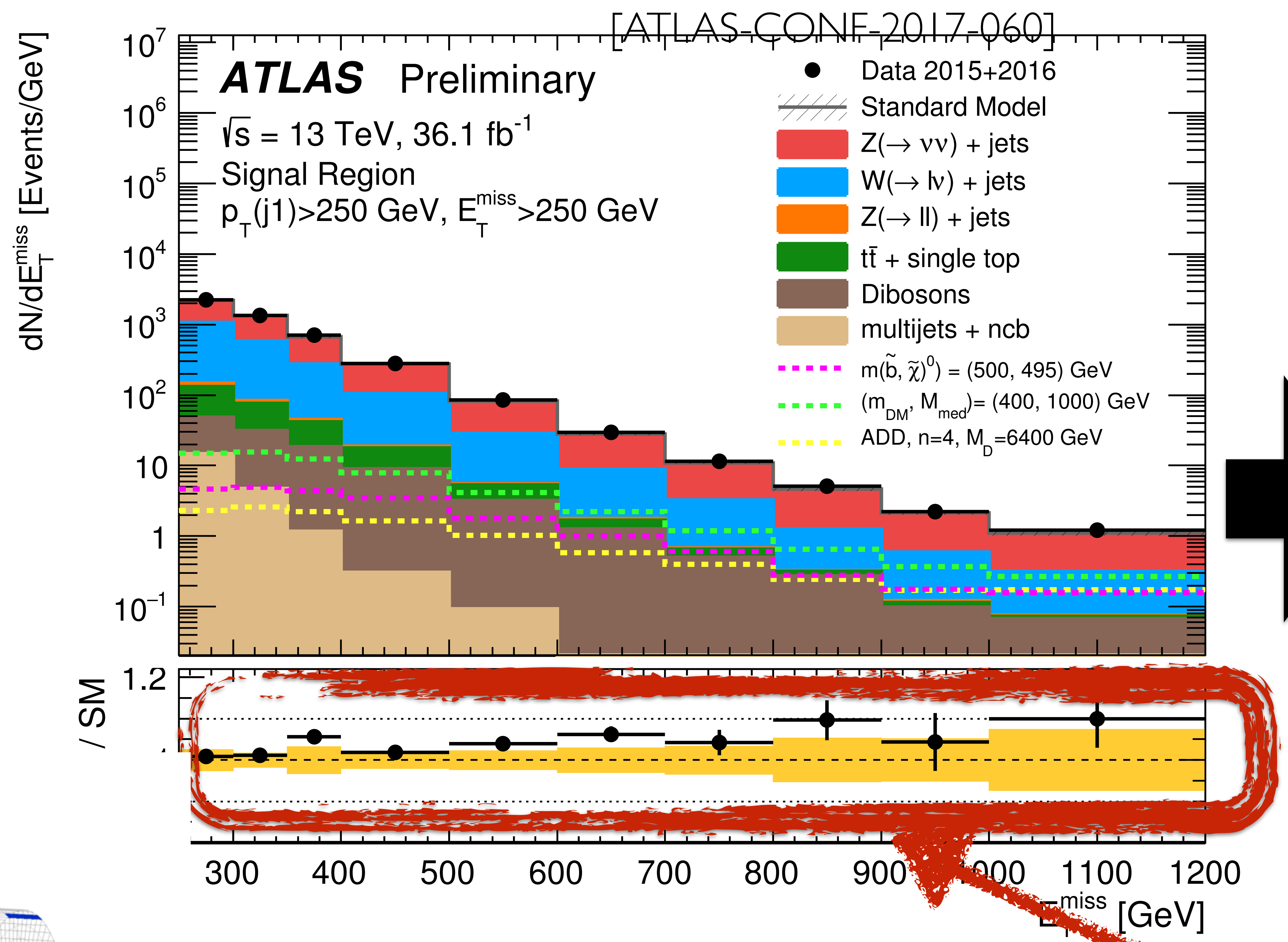
ADD

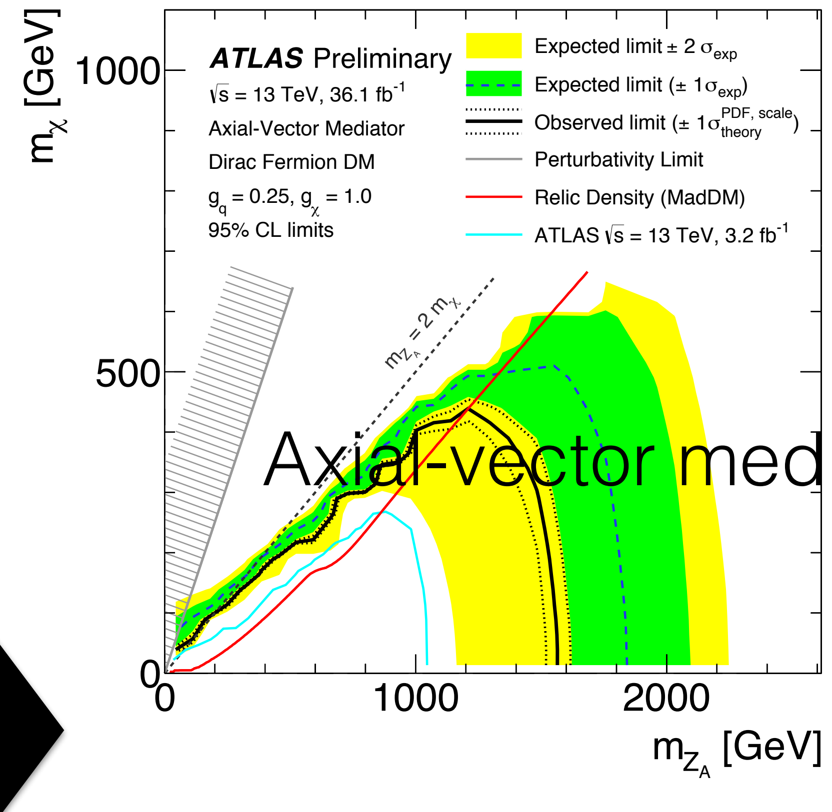
O(5-10%) error @ 1 TeV
"precision search"

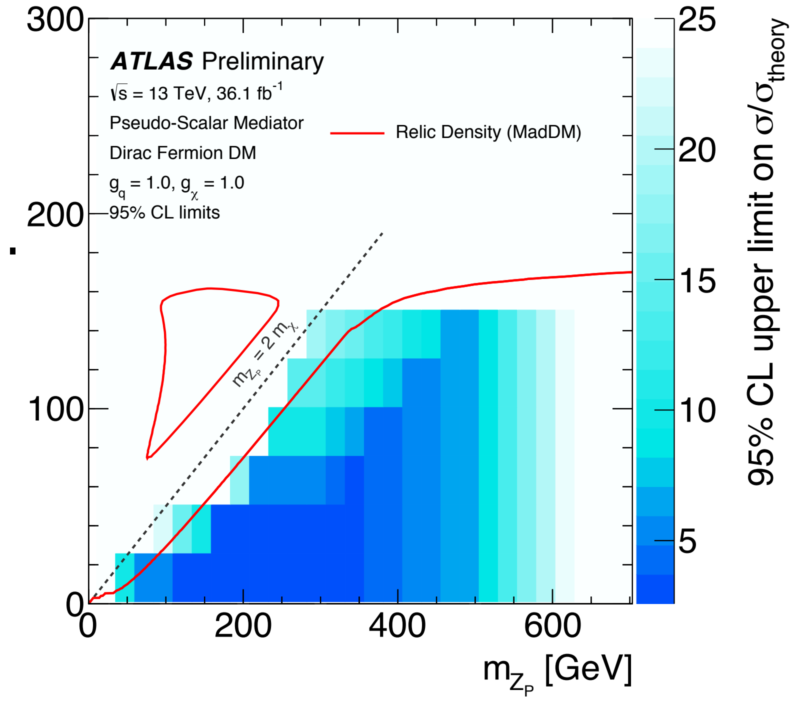
SM backgrounds in monojet / MET+X searches

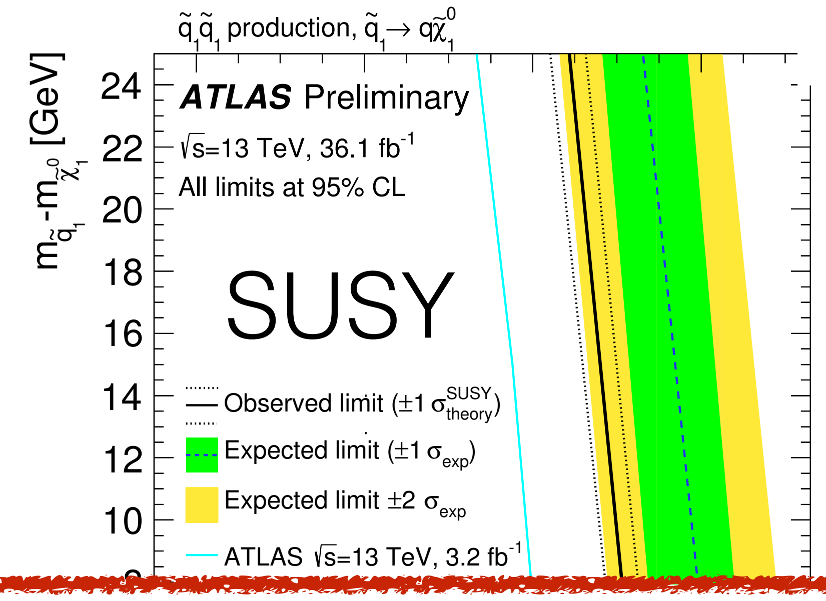
we hope to see:

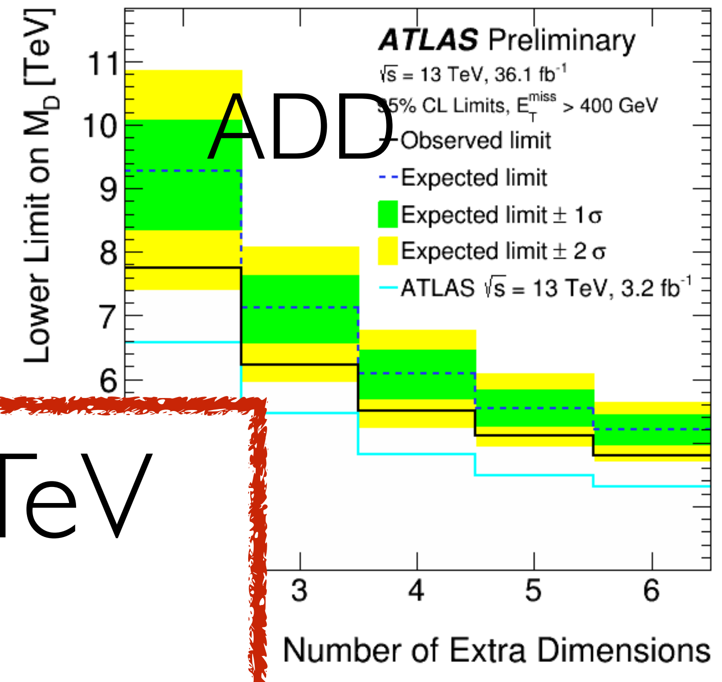




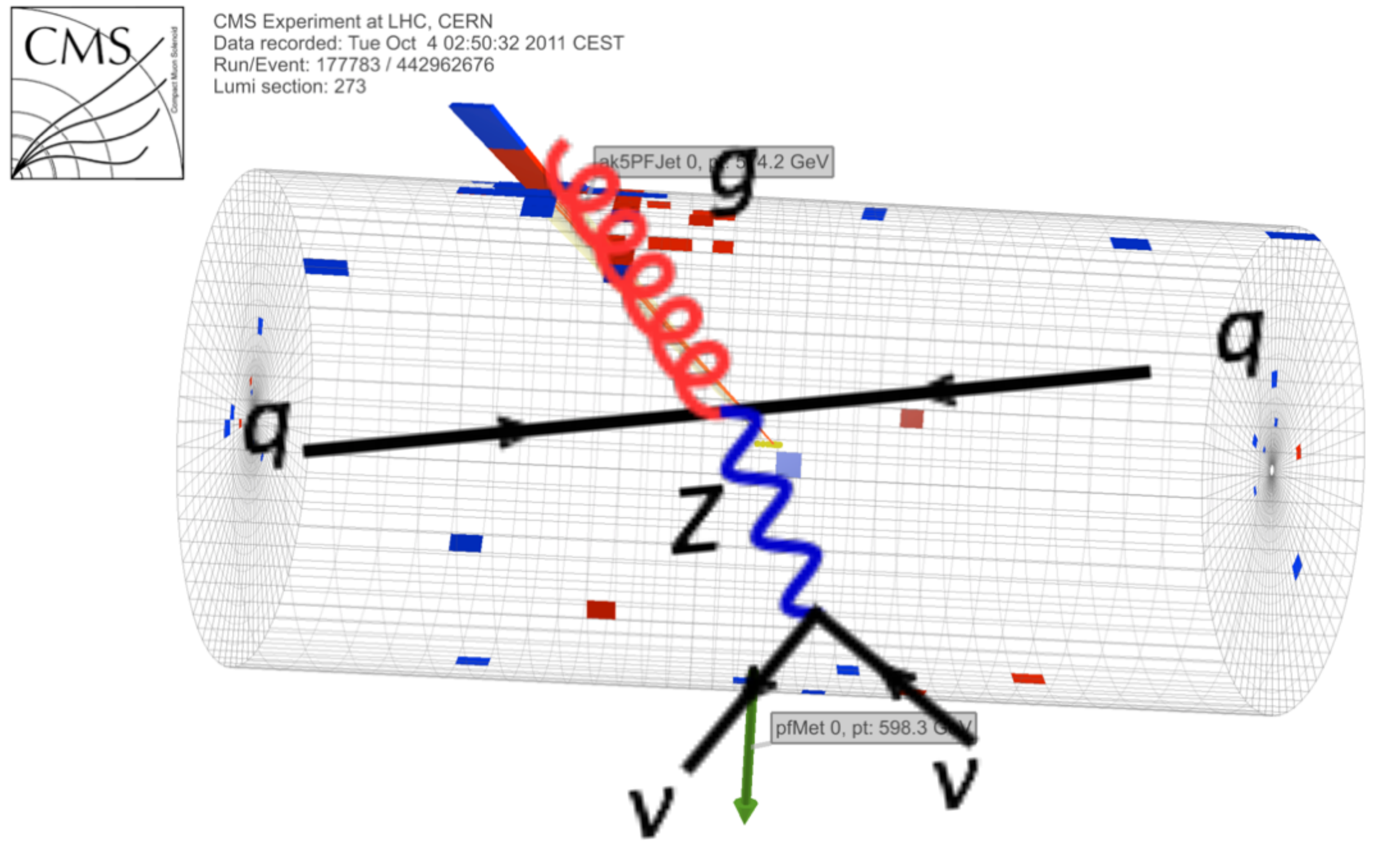








O(5-10%) error @ 1 TeV
"precision search"



missing transverse energy

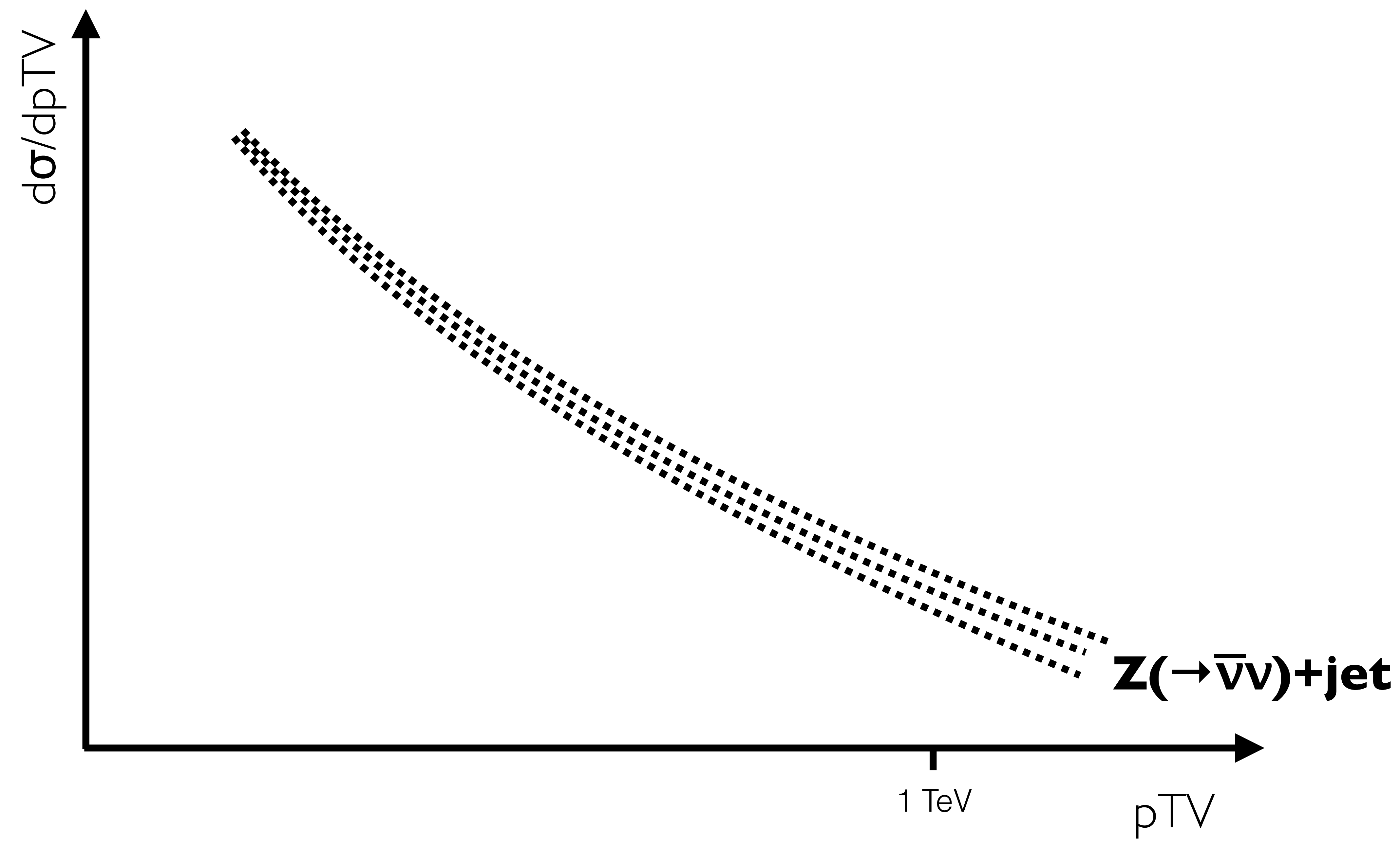
irreducible SM backgrounds:

pp→Z(→$\nu\bar{\nu}$)+jets ⟹ MET + jets

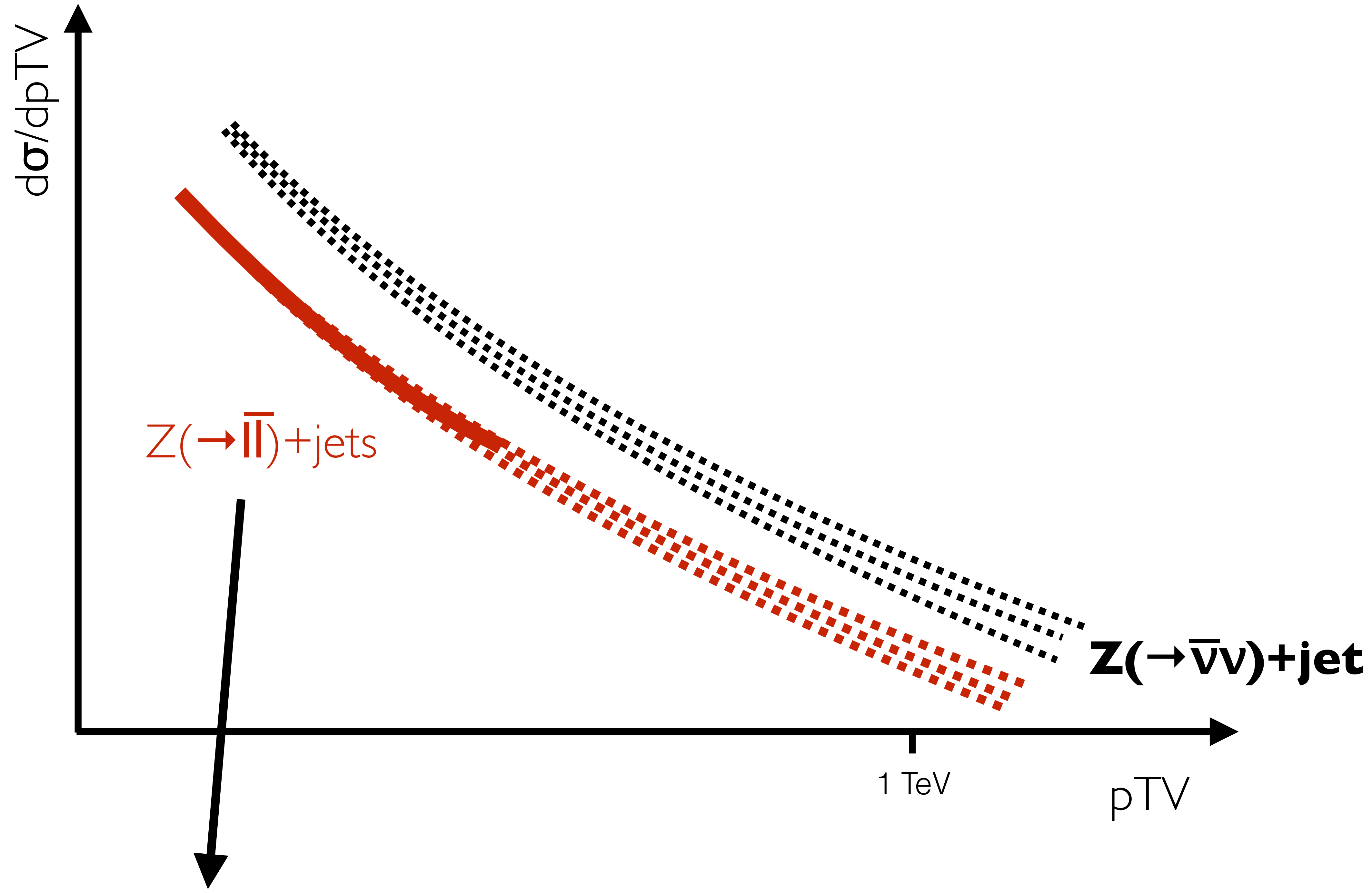
pp→W(→lν)+jets ⟹ MET + jets (lepton lost)

} V + jets

Determine V+jets DM backgrounds

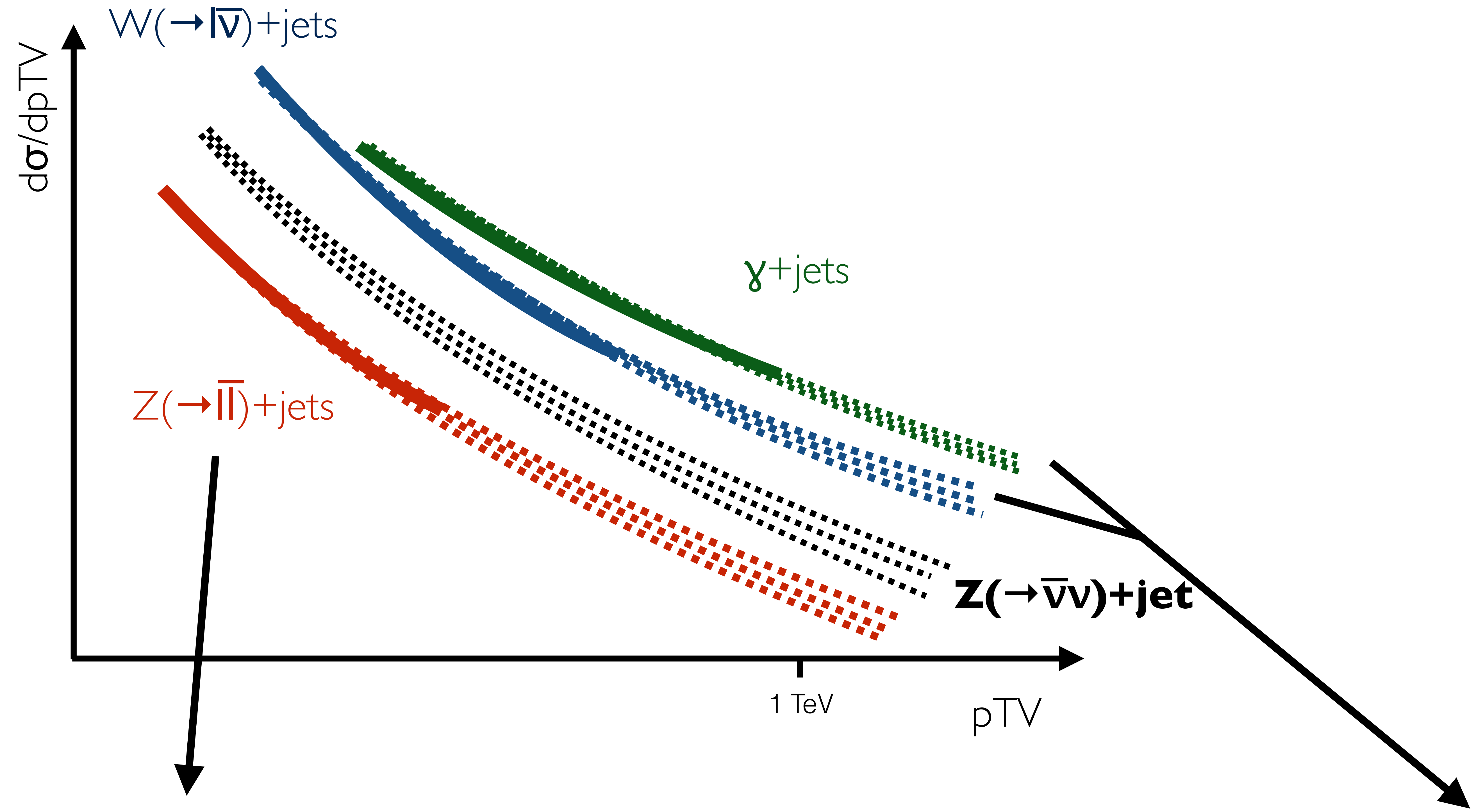


Determine V+jets DM backgrounds



- hardly any systematics (just QED dressing)
- very precise at low pT
- but: limited statistics at large pT

Determine V+jets DM backgrounds



- hardly any systematics (just QED dressing)
- very precise at low pT
- but: limited statistics at large pT

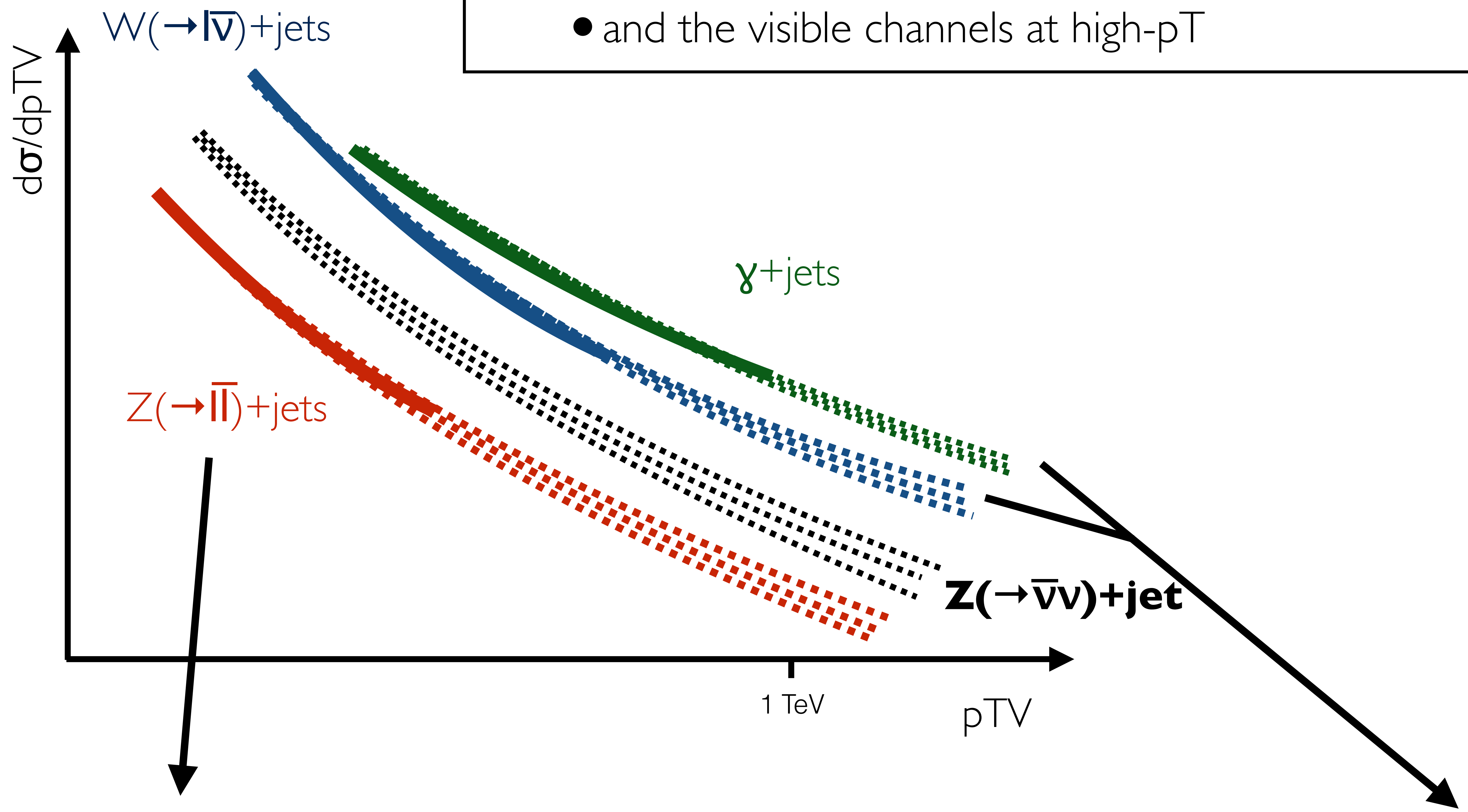
- fairly large data samples at large pT
- systematics from transfer factors

Determine V+jets DM backgrounds

"Normalization to control regions" via **global fit** of

Z(→$l\bar{l}$)+jets, W(→$l\bar{\nu}$)+jets and γ+jets measurements

- to determine Z(→$\bar{\nu}\nu$)+jet
- and the visible channels at high-pT



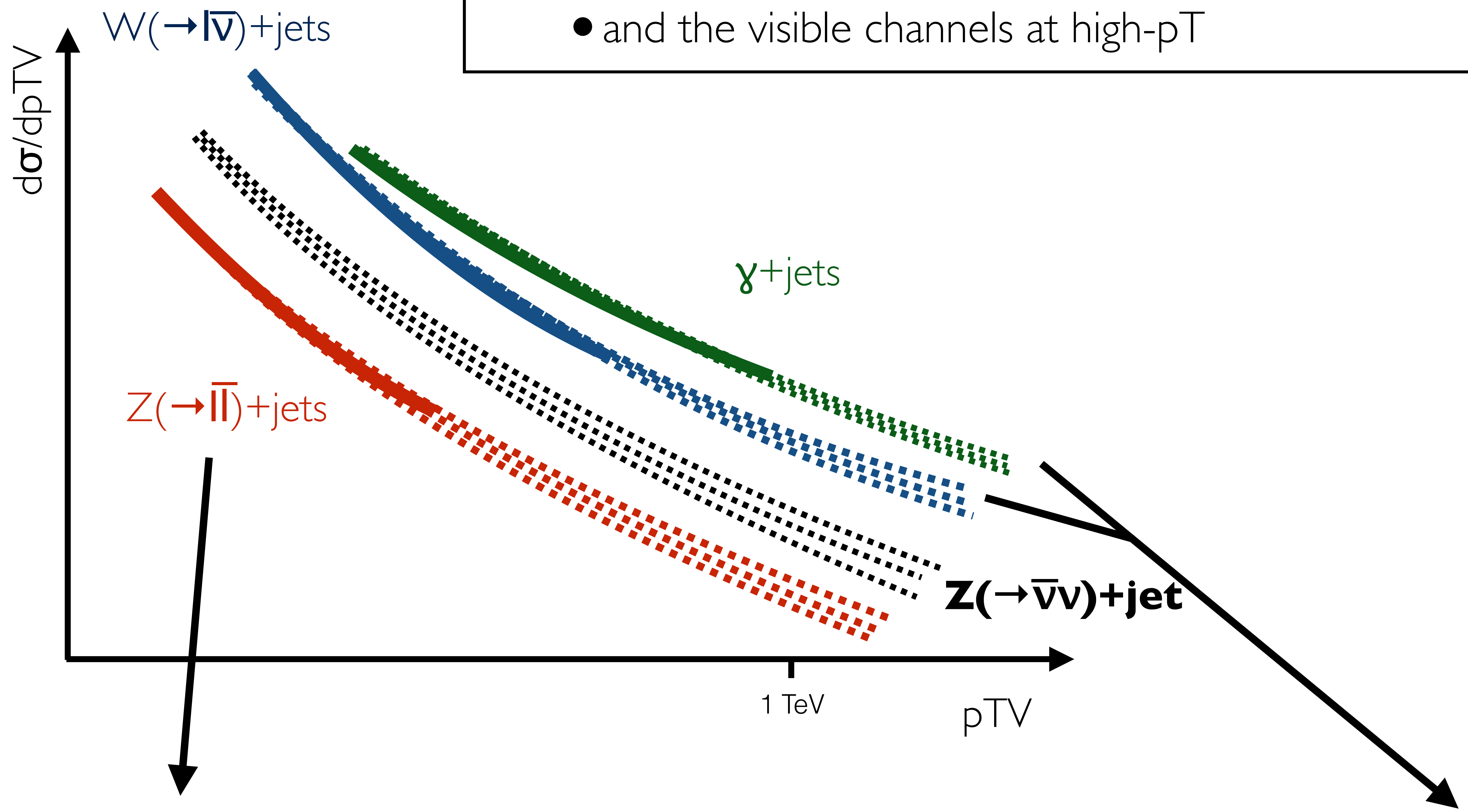
- hardly any systematics (just QED dressing)
- very precise at low pT
- but: limited statistics at large pT

- fairly large data samples at large pT
- systematics from transfer factors

Determine V+jets DM backgrounds

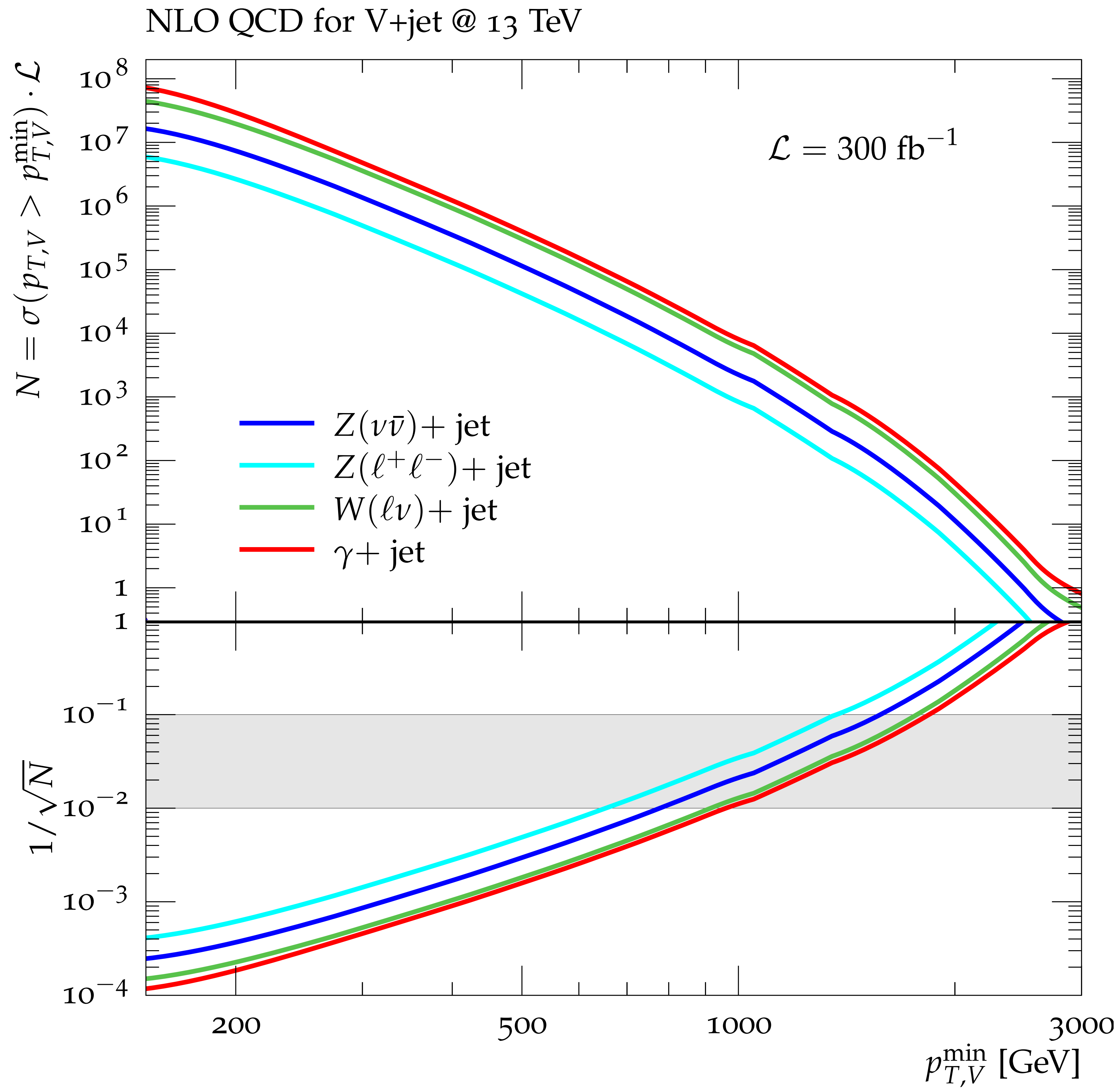
"Normalization to control regions" via **global fit** of
Z(→$l\bar{l}$)+jets, W(→$l\bar{\nu}$)+jets and γ+jets measurements

- to determine Z(→$\bar{\nu}\nu$)+jet
- and the visible channels at high-pT



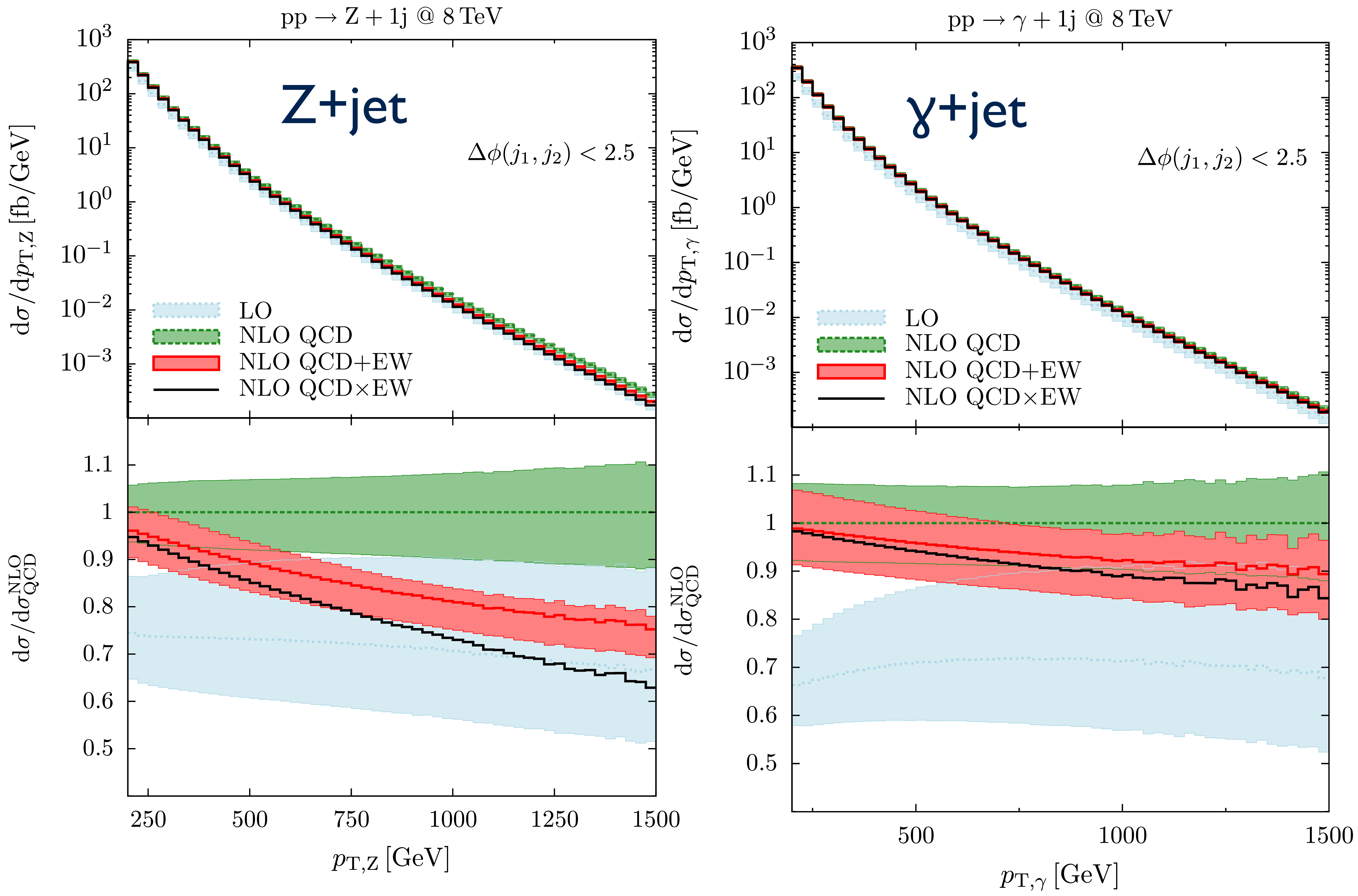
- hardly any systematics (just QED dressing)
- very precise at low pT
- but: limited statistics at large pT

- fairly large data samples at large pT
- systematics from transfer factors



- for 500 GeV < pTV < 1000 GeV: background statistics will be at **1% level**
- this level of precision is theoretically possible @ **NNLO QCD + NNLO EW**

Prelude: Z+jet vs. γ + 1 jet



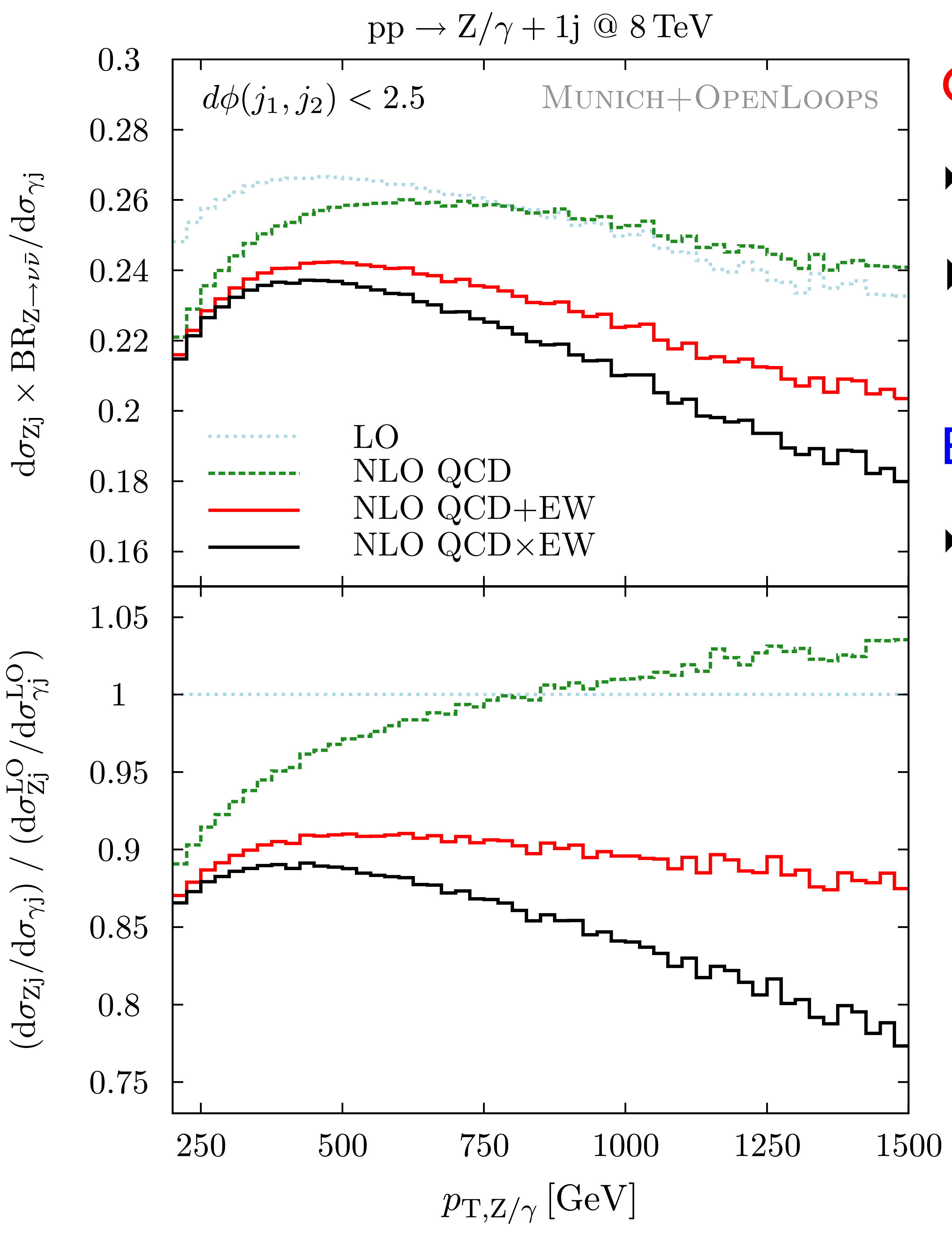
QCD corrections

- mostly moderate and stable QCD corrections
- (almost) **identical QCD corrections in the tail**, sizeable differences for small pT

EW corrections

- **correction in pT(Z) > correction in pT(γ)**
- **-20/-8%** for Z/γ at 1 TeV
- EW corrections > QCD uncertainties for $p_{T,Z}$ > 350 GeV

Prelude: Z/γ pT-ratio



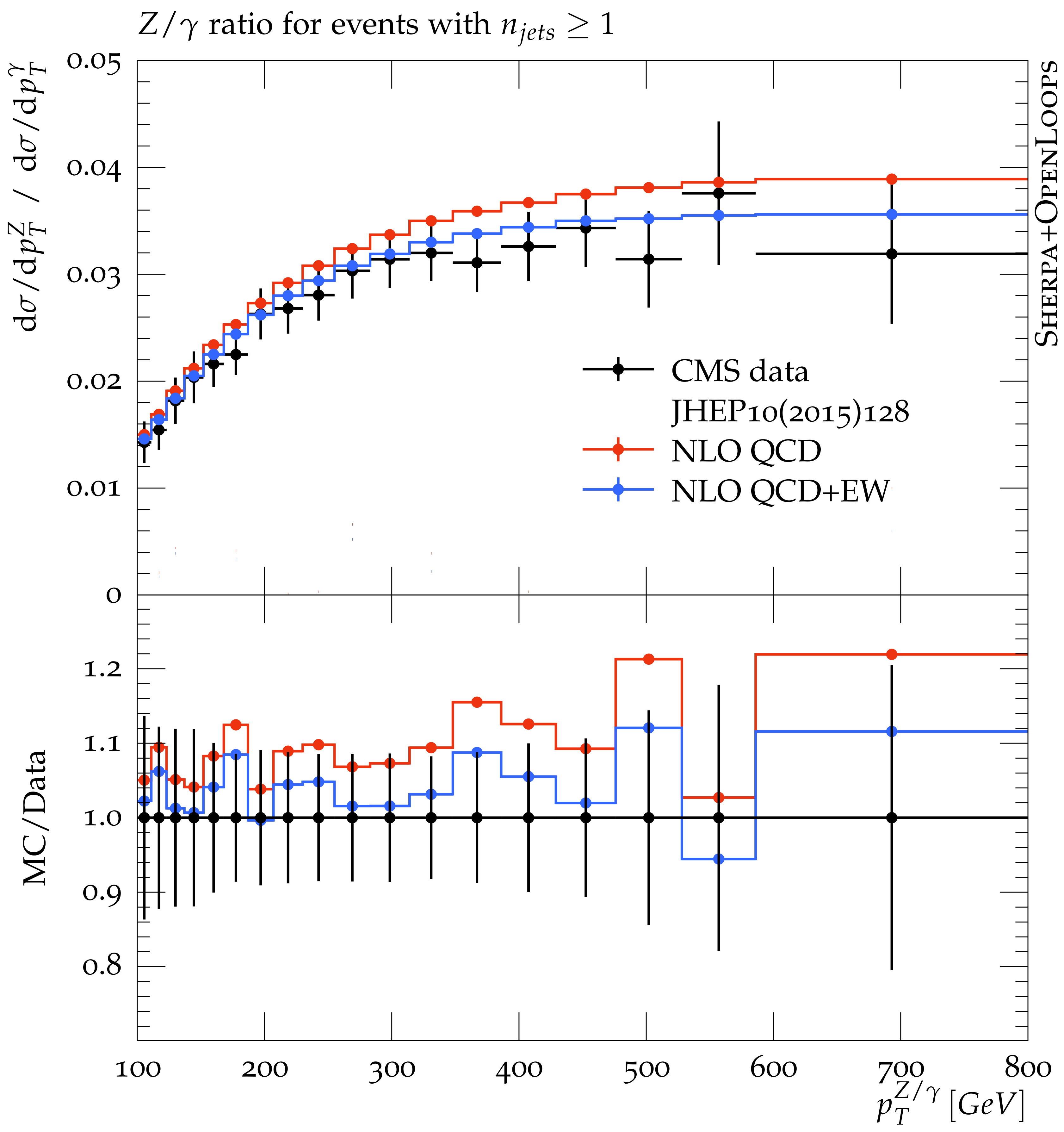
QCD corrections

- 10-15% below 250 GeV
- ≲ 5% above 350 GeV

EW corrections

- sizeable difference in EW corrections results in 10-15% corrections at several hundred GeV

*[Ciulli, Kallweit, JML, Pozzorini, Schönherr for **LesHouches'15**]*



- remarkable agreement with data at @ NLO QCD+EW!

Uncertainty estimates
at
(N)NLO QCD + (n)NLO EW

how to correlate scale uncertainties in ratios?

how to estimate uncertainties due to missing higher-order EW?

how to combine higher-order QCD and EW correction?
what is the related uncertainty?

[1705.04664]

Precise predictions for V+jet DM backgrounds

work in collaboration with:
R. Boughezal, J.M. Campell, A. Denner, S. Dittmaier, A. Huss, A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, S. Kallweit, M. L. Mangano, P. Maierhöfer, T.A. Morgan, A. Mück, M. Schönherr, F. Petriello, S. Pozzorini, G. P. Salam, C.Williams

- Combination of state-of-the-art predictions: (N)NLO QCD+(n)NLO EW in order to match (future) experimental sensitivities (1-10% accuracy in the few hundred GeV-TeV range)

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}\vec{y}}\sigma^{(V)}(\vec{\varepsilon}_{\mathrm{MC}},\vec{\varepsilon}_{\mathrm{TH}}) := \frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}\vec{y}}\sigma^{(V)}_{\mathrm{MC}}(\vec{\varepsilon}_{\mathrm{MC}})\left[\frac{\frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{TH}}(\vec{\varepsilon}_{\mathrm{TH}})}{\frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{MC}}(\vec{\varepsilon}_{\mathrm{MC}})}\right]$$

one-dimensional reweighting of MC samples in $x = p_{\mathrm{T}}^{(V)}$

with $$\frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{TH}} = \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{QCD}} + \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{mix}} + \frac{\mathrm{d}}{\mathrm{d}x}\Delta\sigma^{(V)}_{\mathrm{EW}} + \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\gamma-\mathrm{ind.}}$$

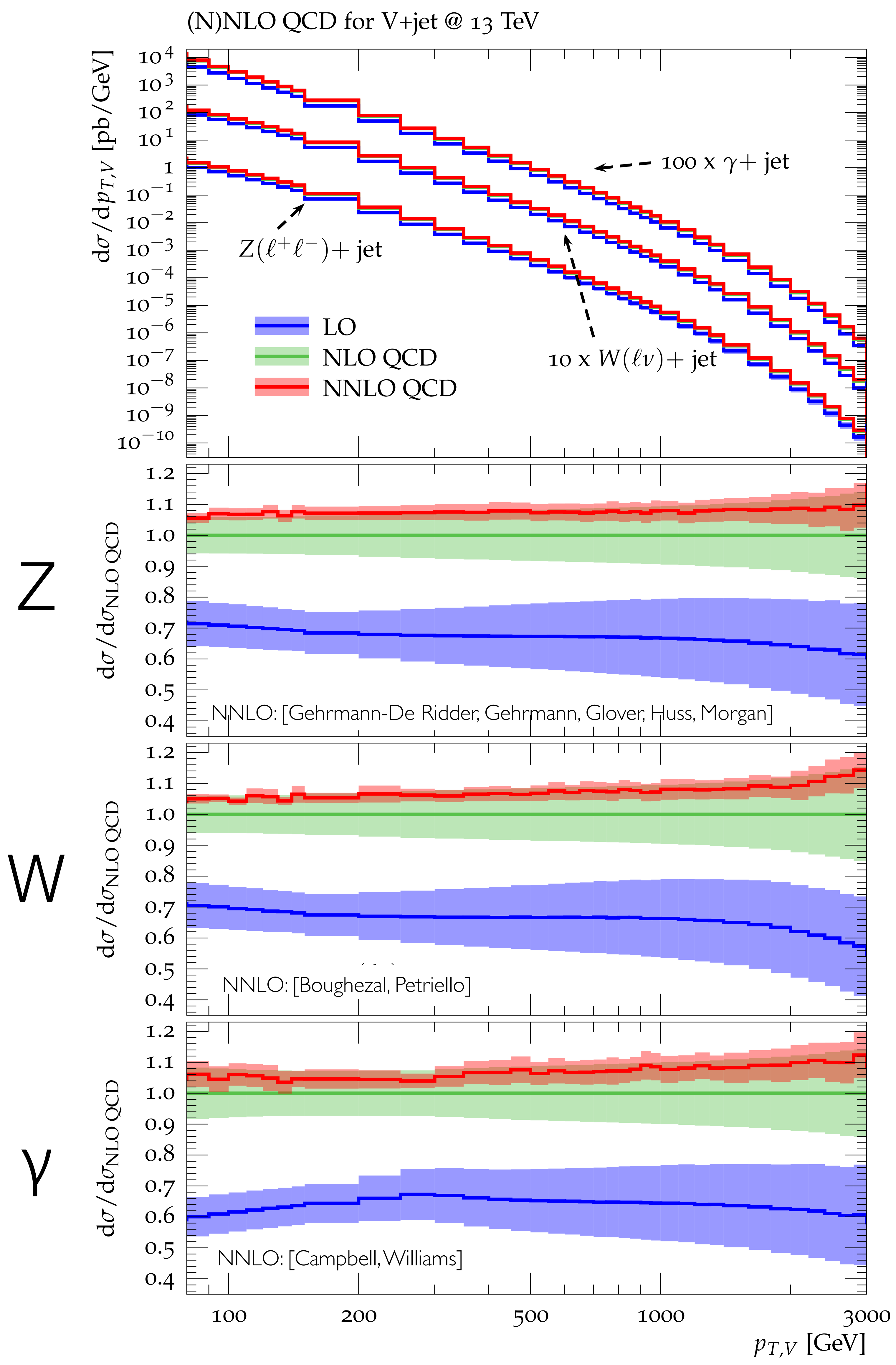
Note: analysis cuts can be considerable different from reweighting setup

- **Robust uncertainty estimates including**
  1. **Pure QCD uncertainties**
  2. **Pure EW uncertainties**
  3. **Mixed QCD-EW uncertainties**
  4. **PDF, γ-induced uncertainties**

- Prescription for **correlation** of these uncertainties
  - within a process (between low-pT and high-pT)
  - across processes

Pure QCD uncertainties

[JML et. al.: 1705.04664]



$$\frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{QCD}} = \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{LO\,QCD}} + \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{NLO\,QCD}} + \frac{\mathrm{d}}{\mathrm{d}x}\sigma^{(V)}_{\mathrm{NNLO\,QCD}}$$

$$\mu_0 = \frac{1}{2}\left(\sqrt{p^2_{\mathrm{T},\ell^+\ell^-} + m^2_{\ell^+\ell^-}} + \sum_{i\in\{q,g,\gamma\}} |p_{\mathrm{T,i}}|\right)$$

this is a 'good' scale for V+jets

- at large pTV: HT'/2 ≈ pTV
- modest higher-order corrections
- sufficient convergence

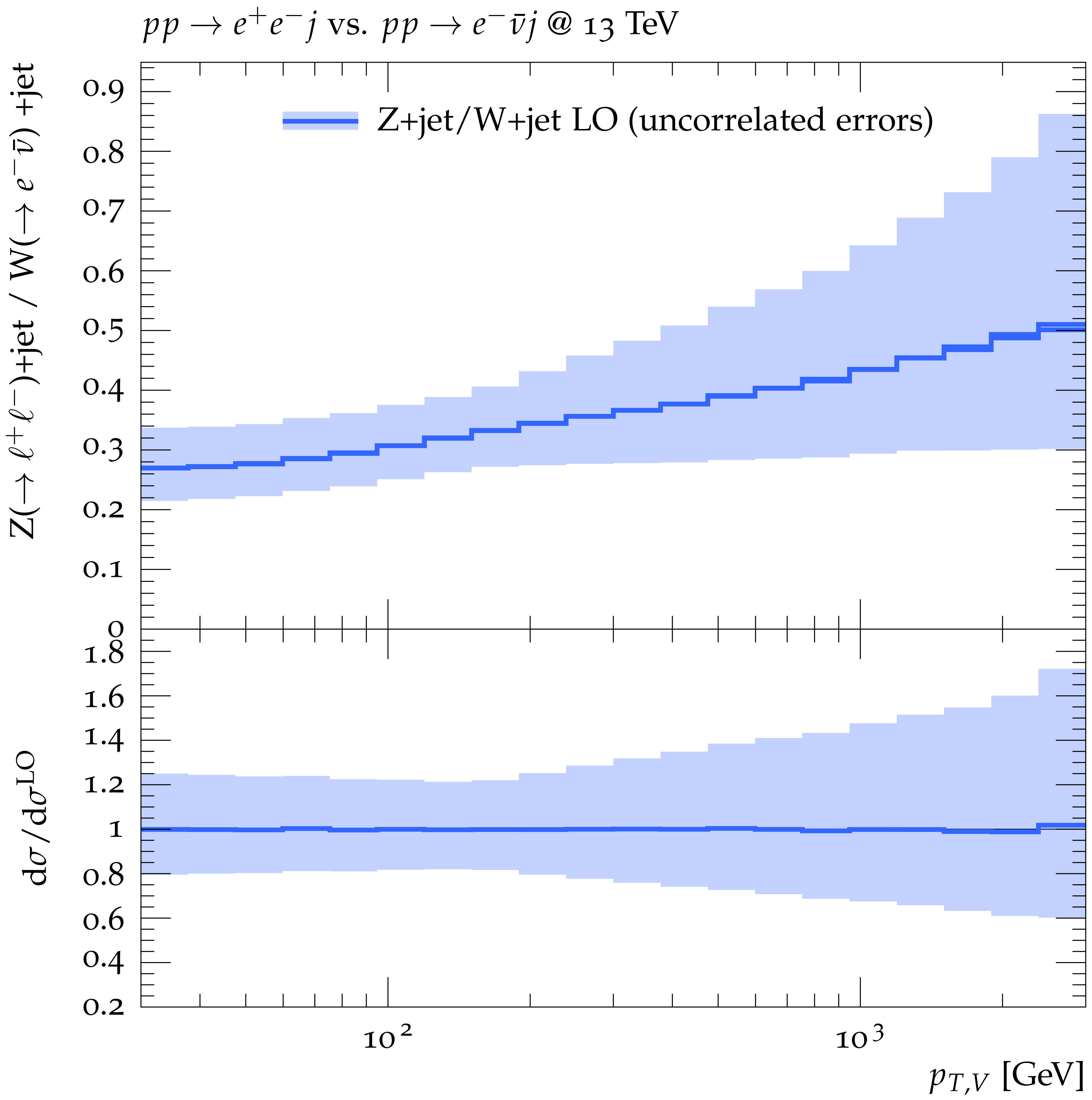
scale uncertainties due to 7-pt variations:

O(20%) uncertainties at LO
O(10%) uncertainties at NLO
O(5%) uncertainties at NNLO

with minor shape variations

How to correlate these uncertainties across processes?

How to correlate QCD uncertainties across processes?

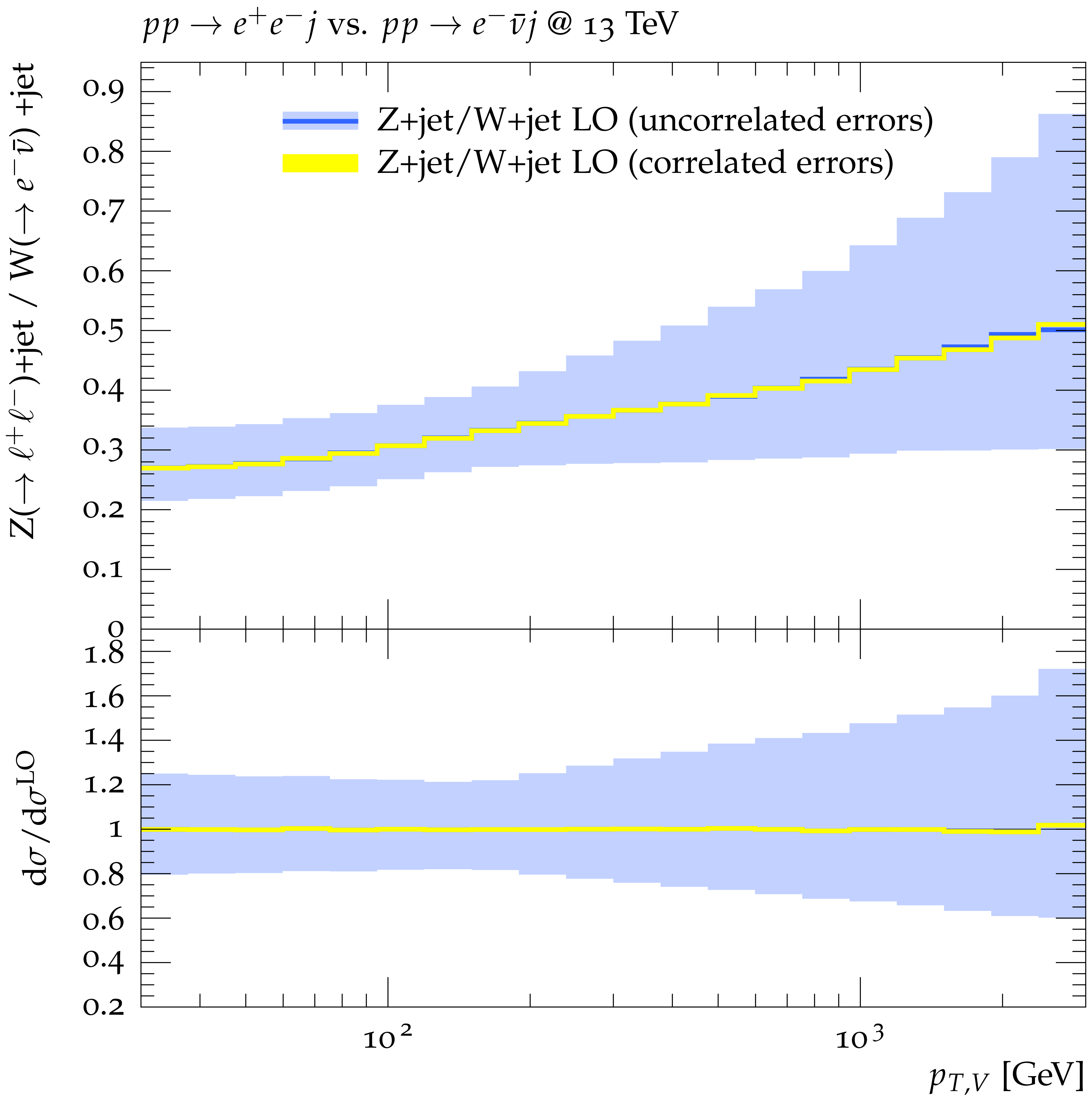


consider Z+jet / W+jet $p_{T,V}$-ratio @ LO

uncorrelated treatment yields O(40%) uncertainties

How to correlate QCD uncertainties across processes?

[1705.04664]



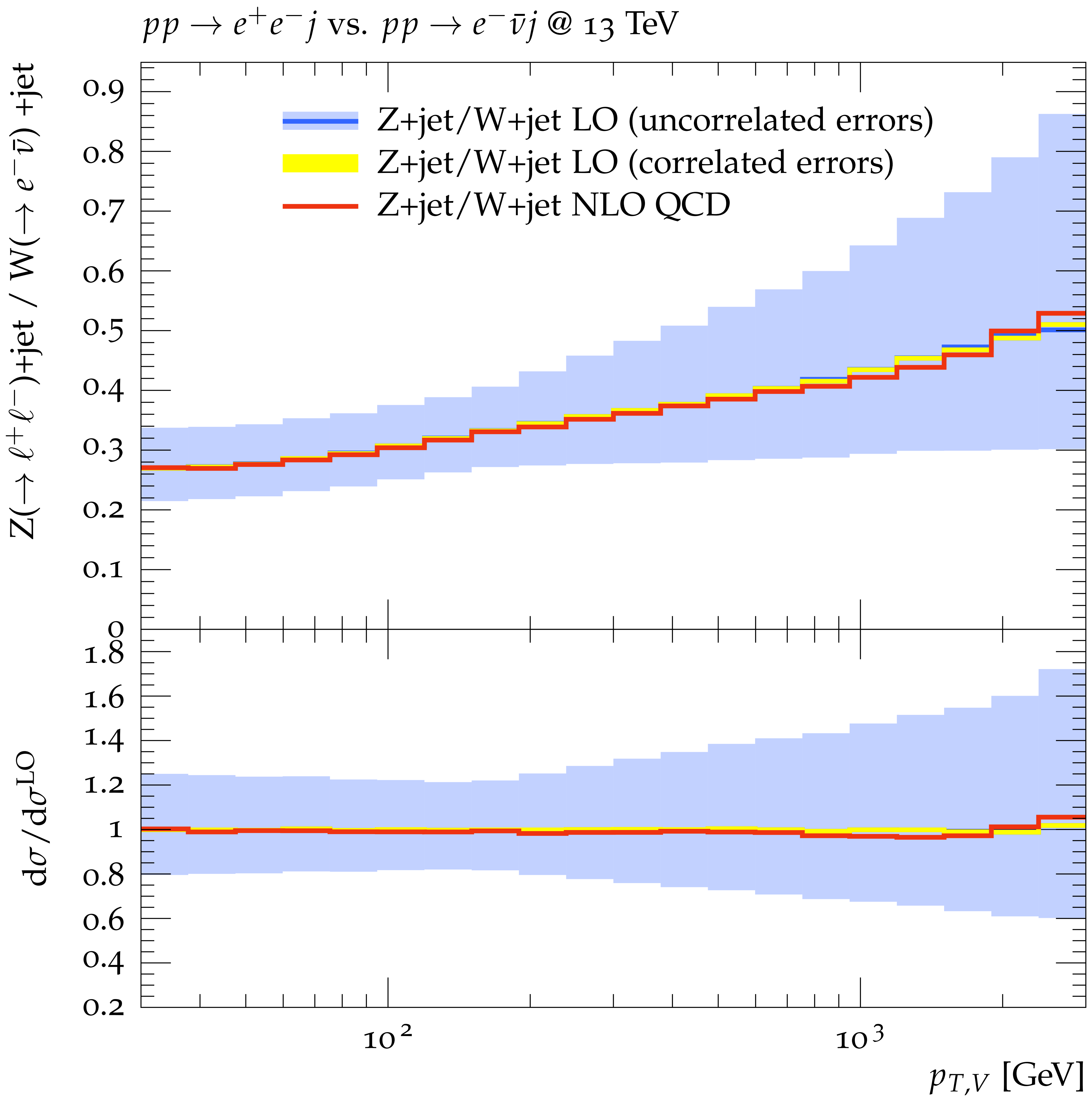
consider Z+jet / W+jet $p_{T,V}$-ratio @ LO

uncorrelated treatment yields O(40%) uncertainties

correlated treatment yields tiny O(<~ 1%) uncertainties

How to correlate QCD uncertainties across processes?

[1705.04664]



consider Z+jet / W+jet $p_{T,V}$-ratio @ LO

uncorrelated treatment yields O(40%) uncertainties

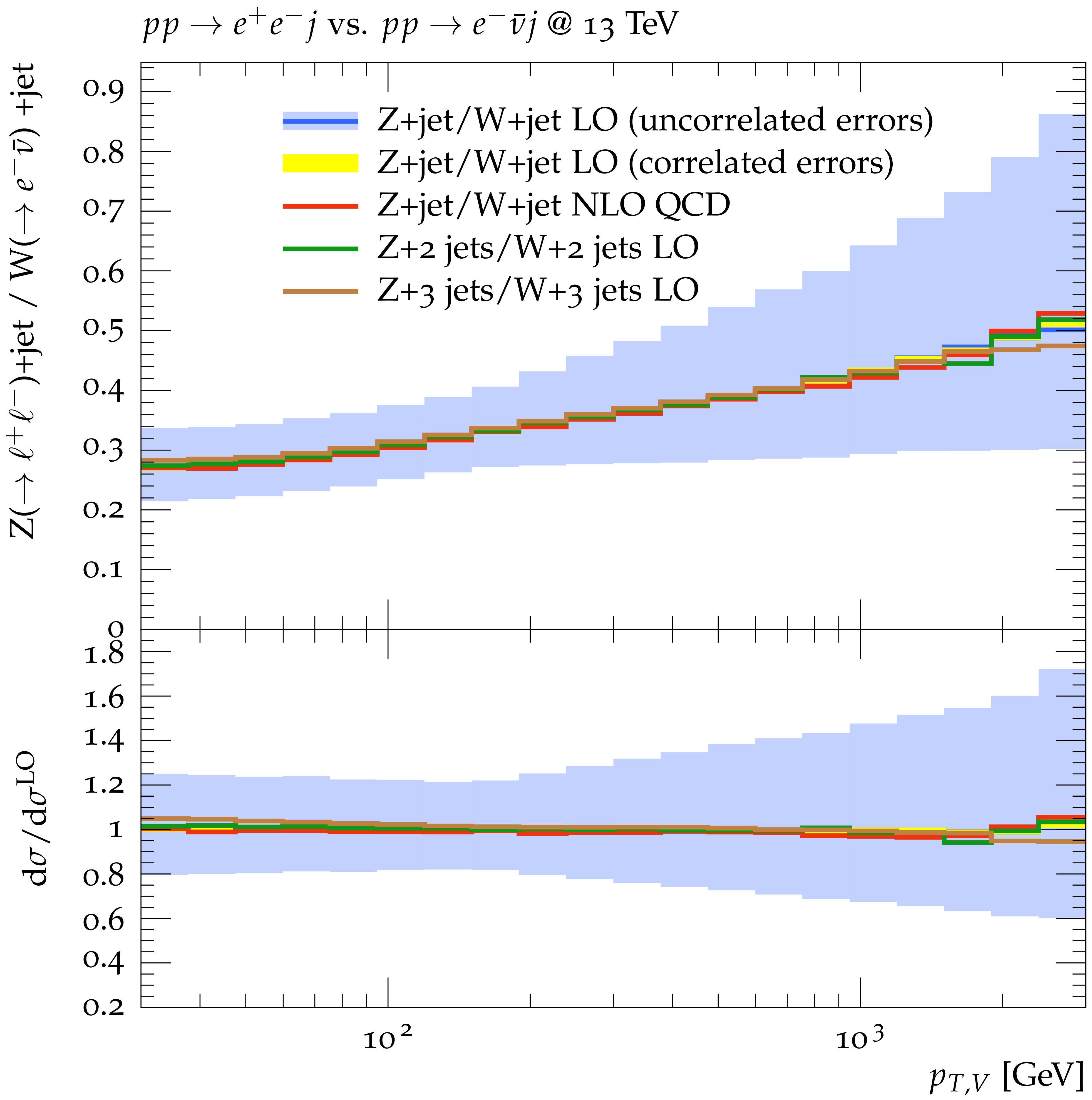
correlated treatment yields tiny O(<~ 1%) uncertainties

check against NLO QCD!

NLO QCD corrections remarkably flat in Z+jet / W+jet ratio!
→ supports correlated treatment of uncertainties!

How to correlate QCD uncertainties across processes?

[1705.04664]



consider Z+jet / W+jet $p_{T,V}$-ratio @ LO

uncorrelated treatment yields O(40%) uncertainties

correlated treatment yields tiny O(<~ 1%) uncertainties

check against NLO QCD!

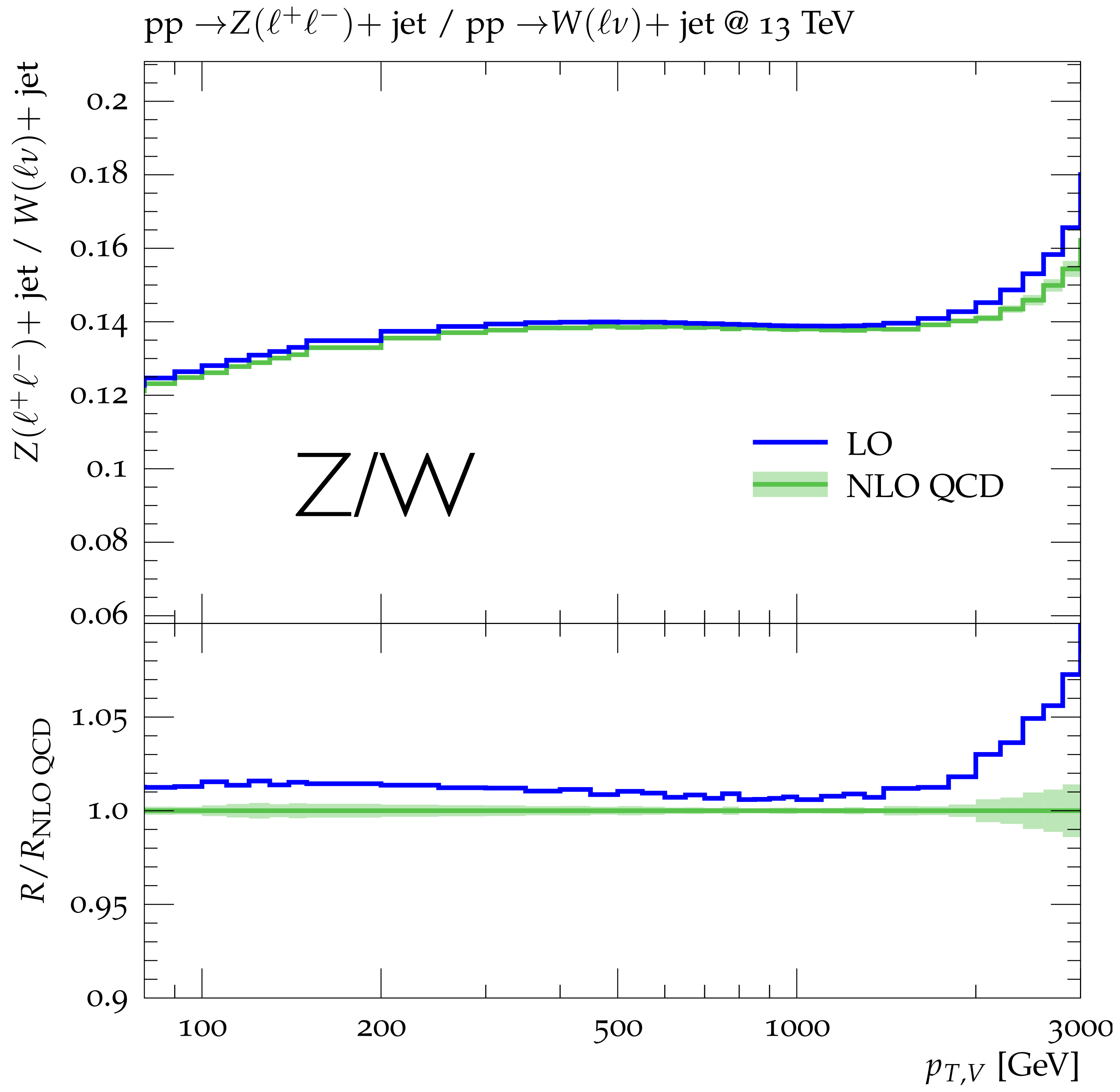
NLO QCD corrections remarkably flat in Z+jet / W+jet ratio!
→ supports correlated treatment of uncertainties!

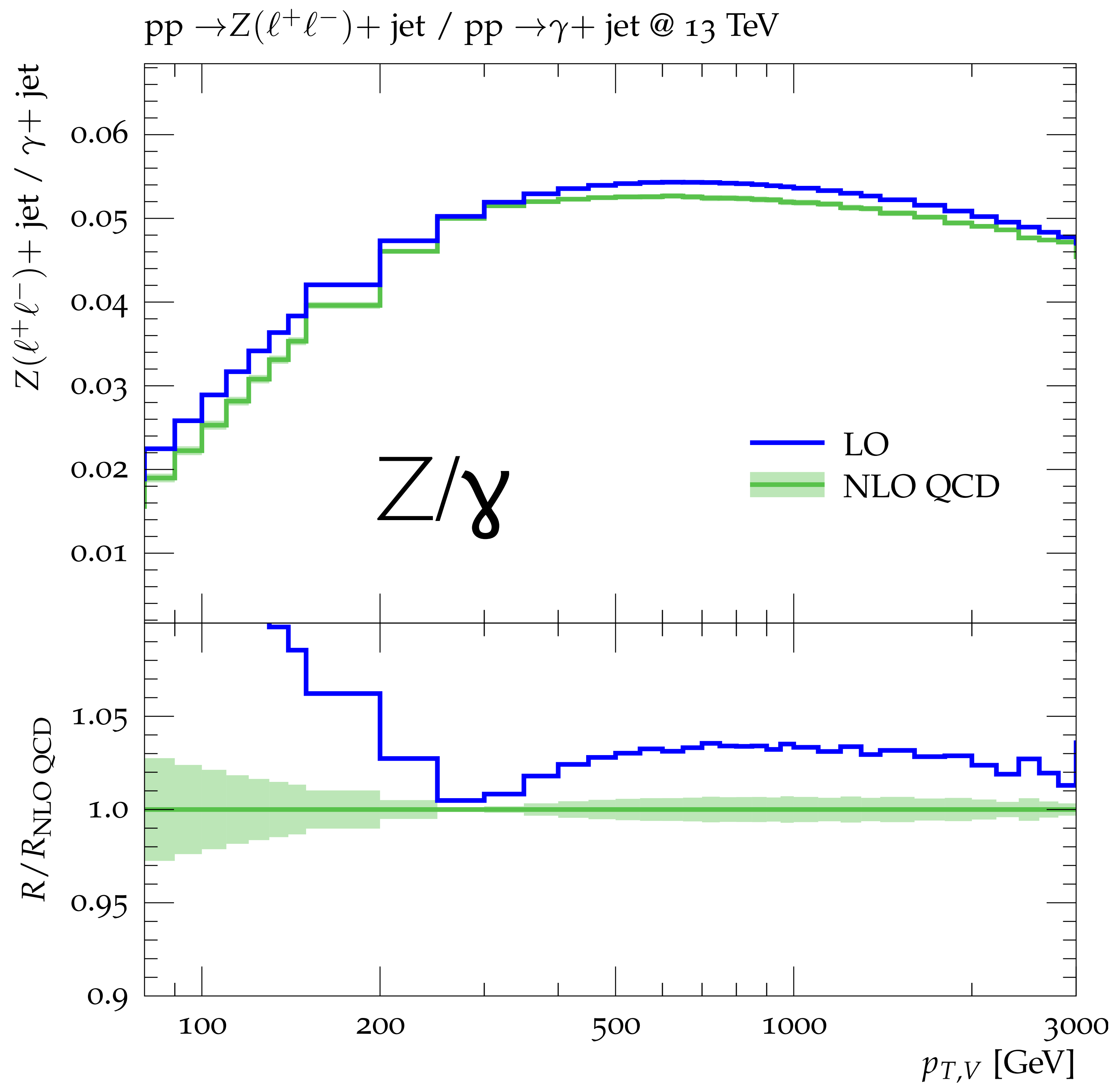
Also holds for higher jet-multiplicities
→ indication of correlation also in higher-order corrections beyond NLO!

QCD uncertainties: ratios

How to correlate these uncertainties across processes?

- take scale uncertainties as fully correlated:
  NLO QCD uncertainties cancel at the <~ 1 % level

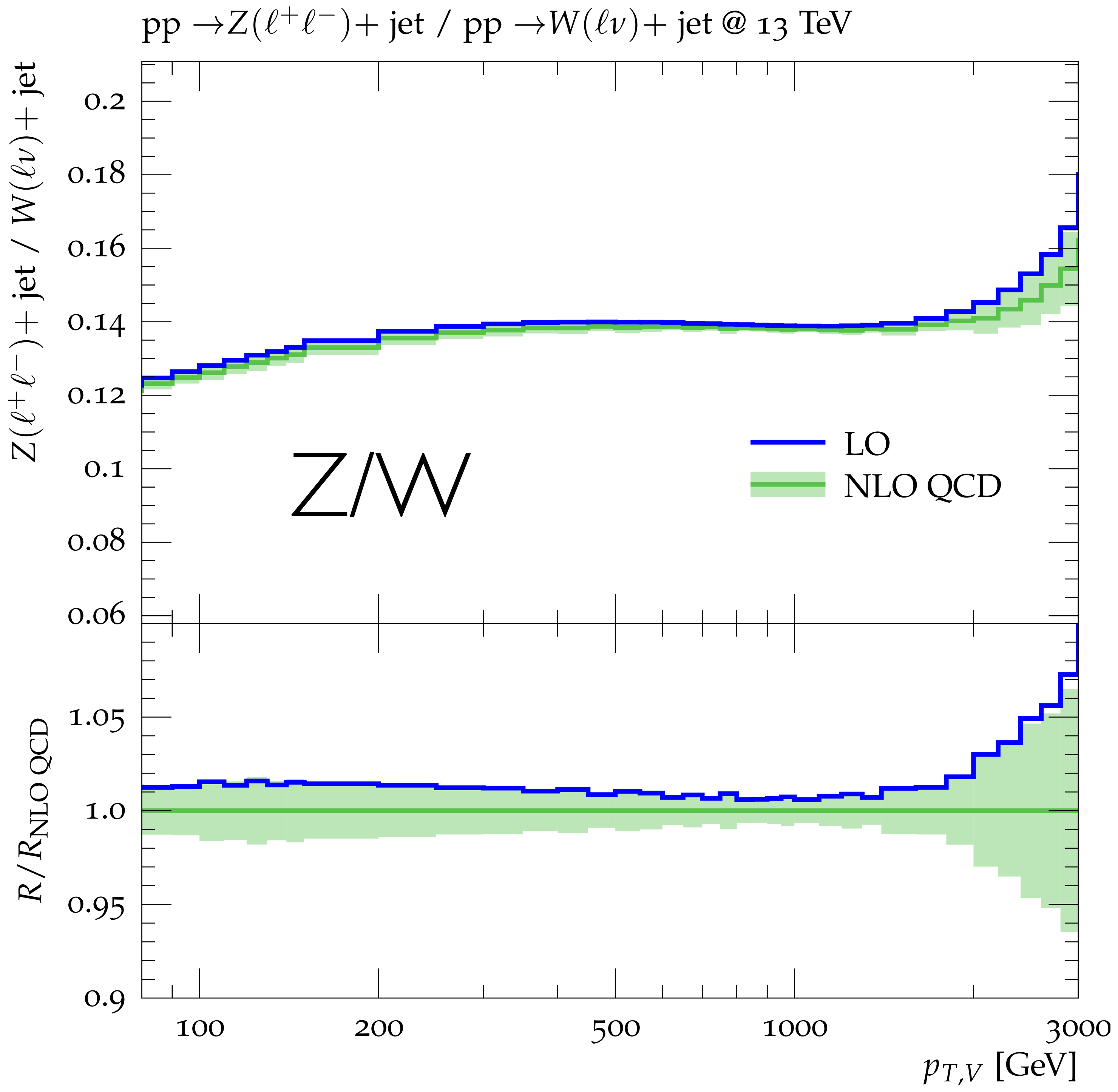




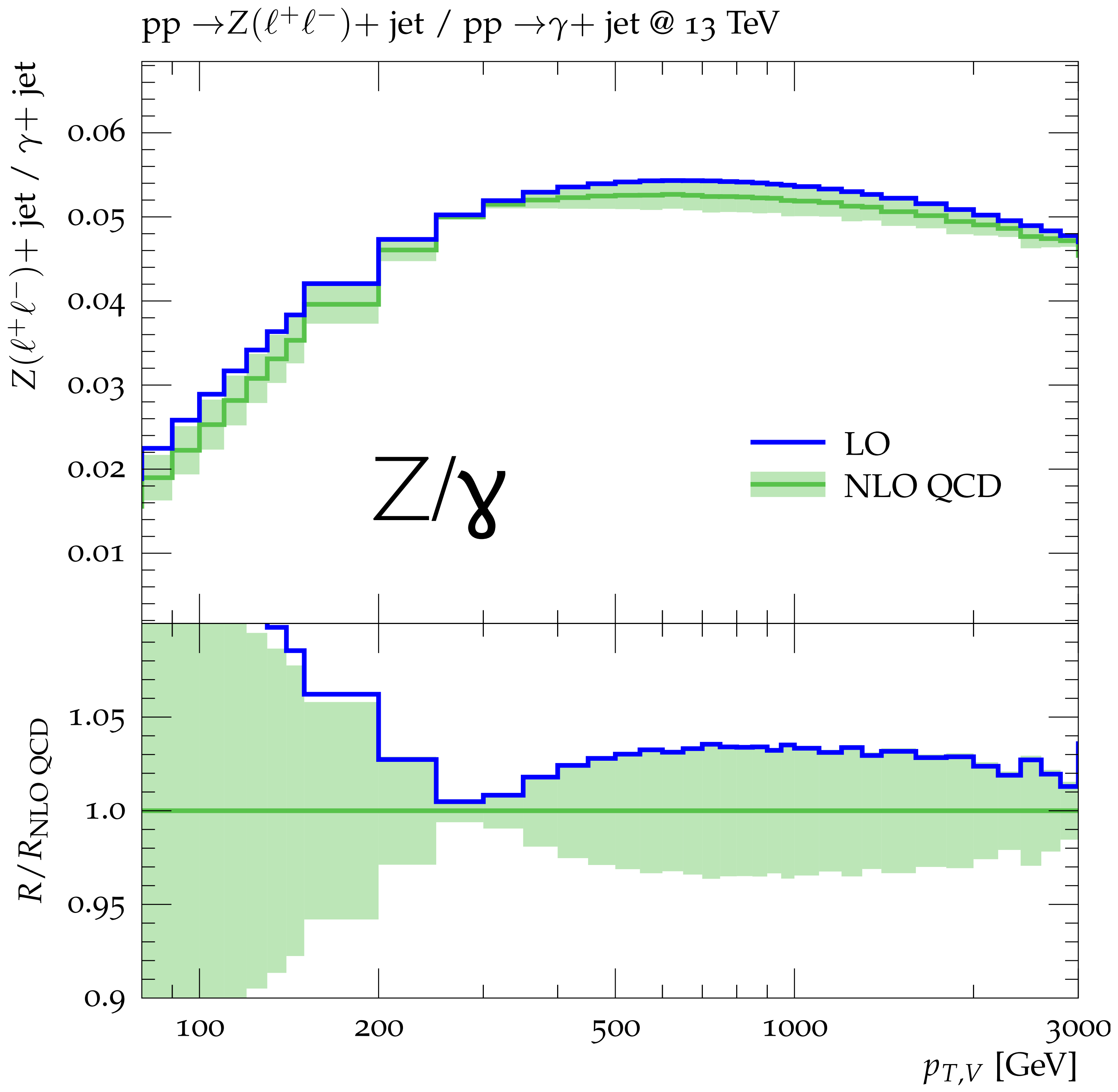
QCD uncertainties: ratios

How to correlate these uncertainties across processes?

- take scale uncertainties as fully correlated:
  NLO QCD uncertainties cancel at the <~ 1 % level
- introduce **process correlation uncertainty** based on K-factor difference: $\delta K_{\mathrm{NLO}} = K^{V}_{\mathrm{NLO}} - K^{Z}_{\mathrm{NLO}}$
  →effectively degrades precision of last calculated order



δ < 2 %

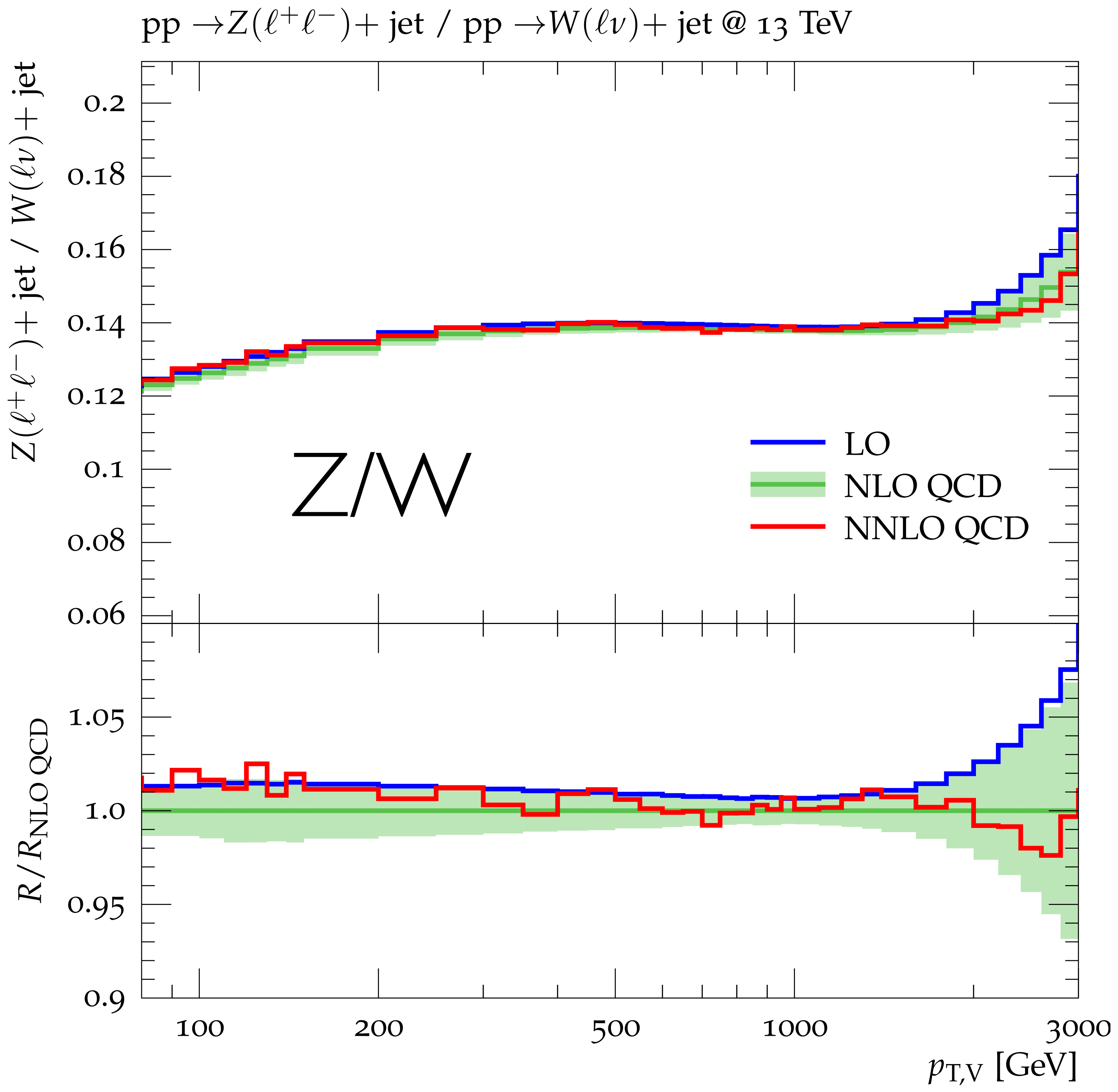


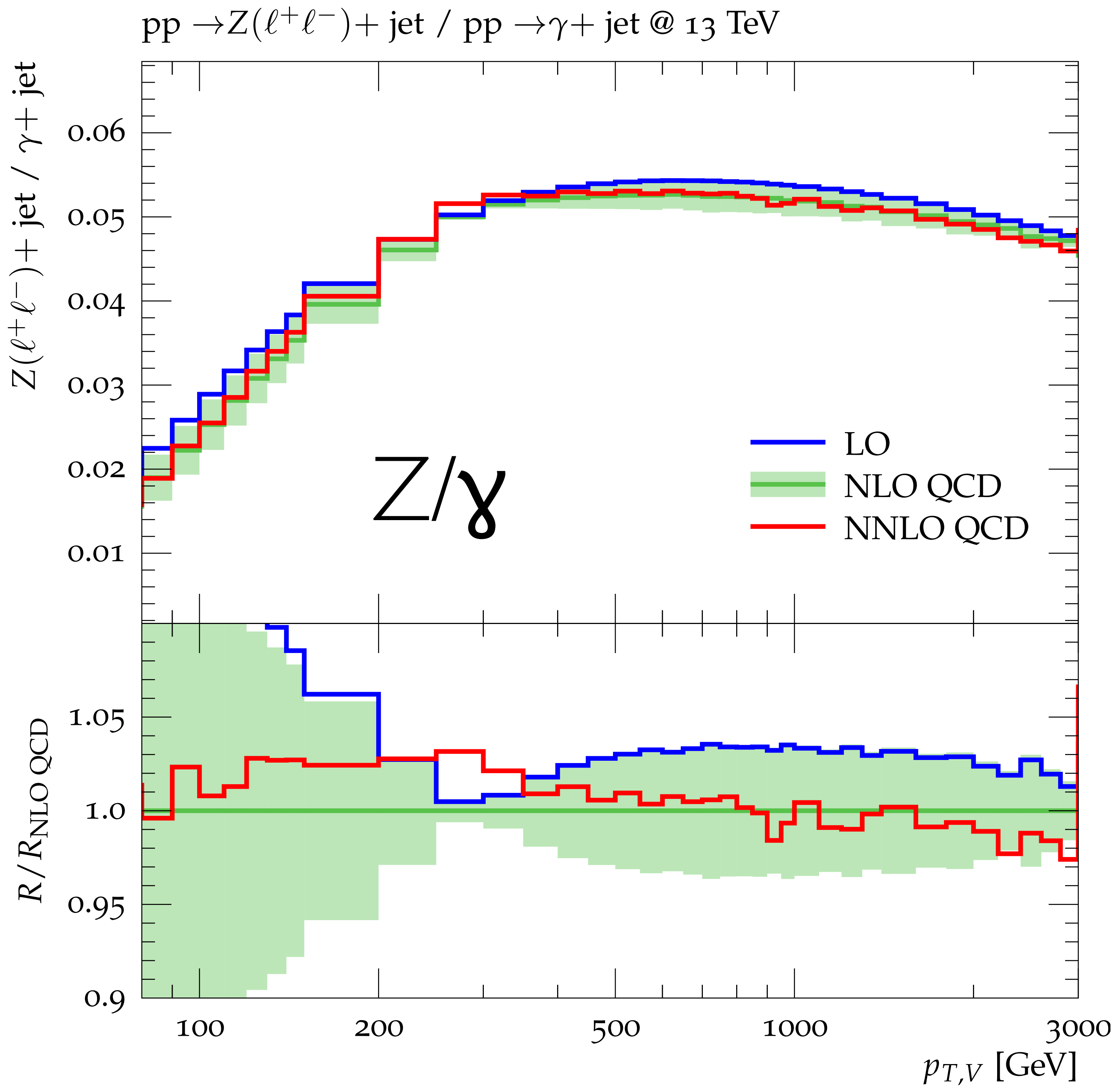
δ < 3-4 %

QCD uncertainties: ratios

How to correlate these uncertainties across processes?

- take scale uncertainties as fully correlated:
  NLO QCD uncertainties cancel at the <~ 1 % level
- introduce **process correlation uncertainty** based on K-factor difference: $\delta K_{\mathrm{NLO}} = K^{V}_{\mathrm{NLO}} - K^{Z}_{\mathrm{NLO}}$
  →effectively degrades precision of last calculated order



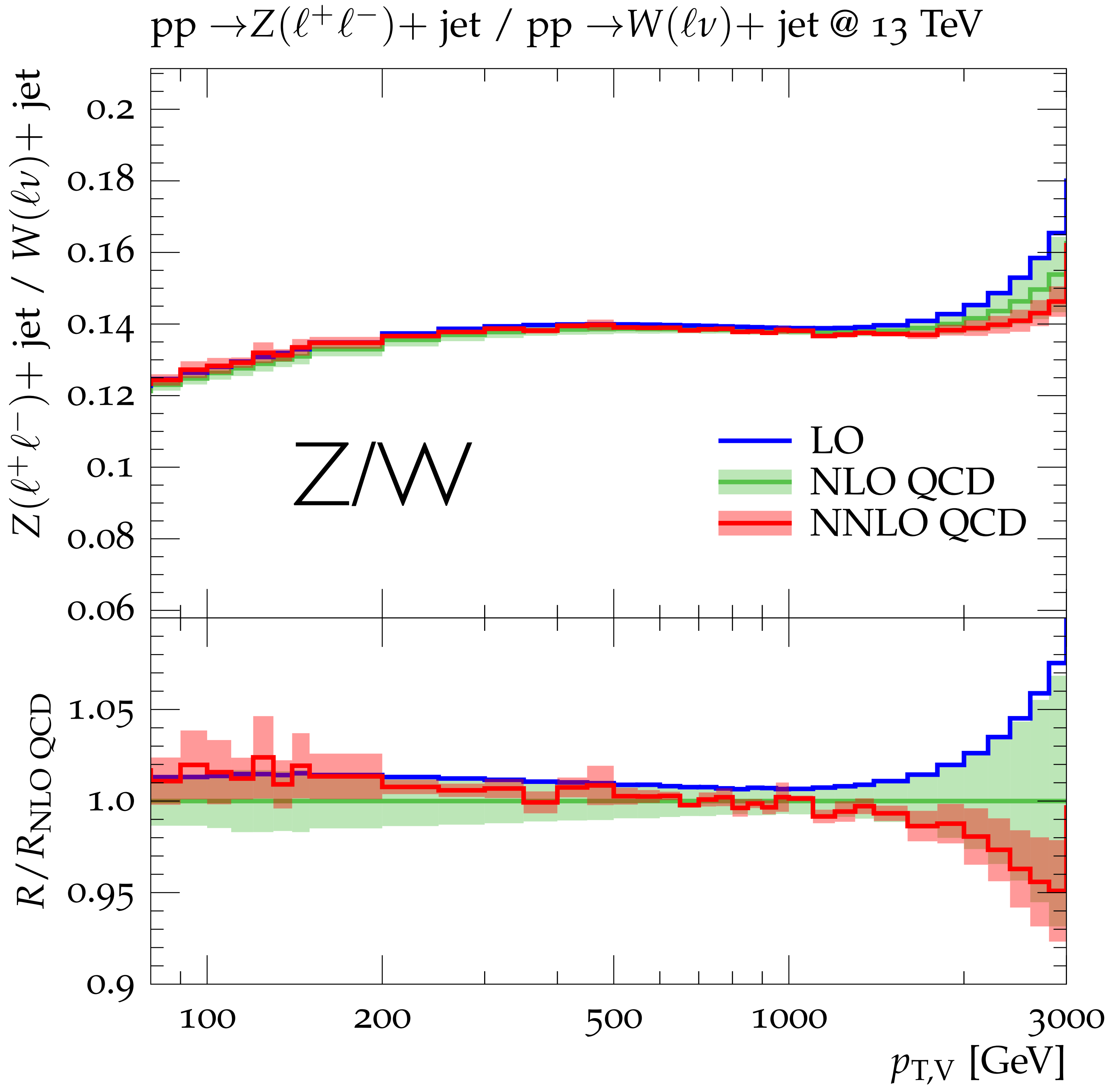


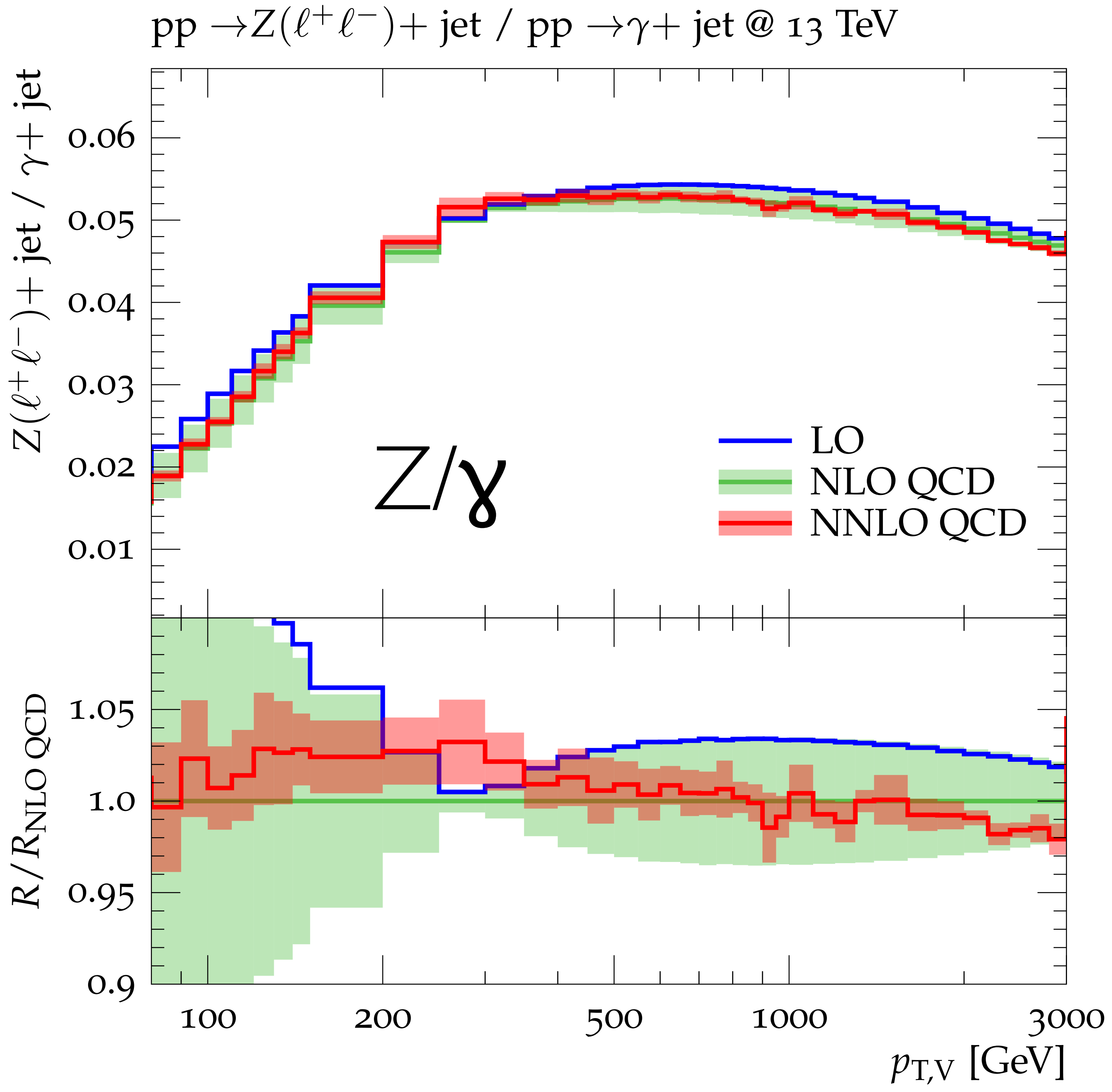
check against NNLO QCD!

QCD uncertainties: ratios

How to correlate these uncertainties across processes?

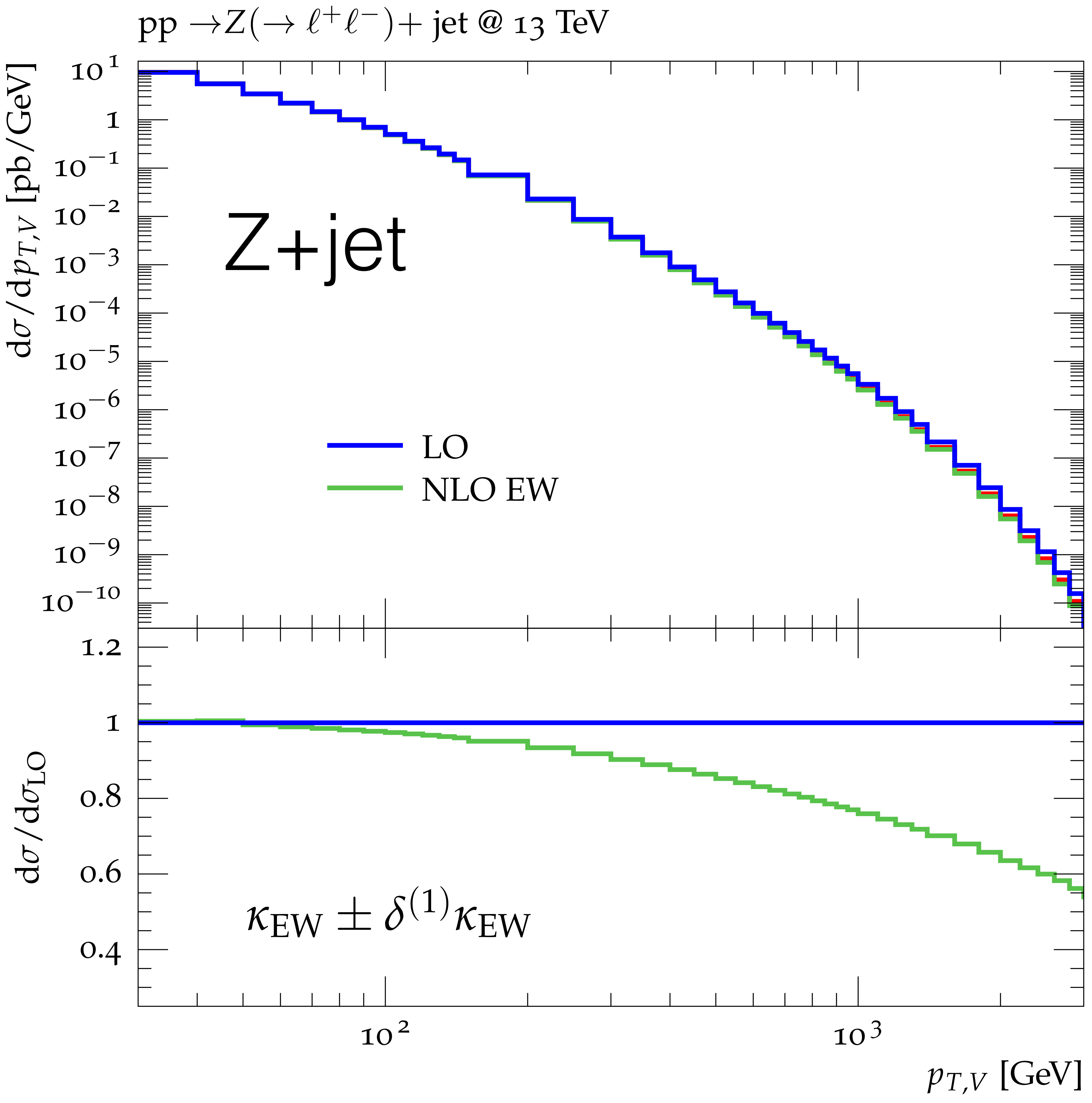
- take scale uncertainties as fully correlated:
  NLO QCD uncertainties cancel at the <~ 1 % level
- introduce **process correlation uncertainty** based on K-factor difference: $\delta K_{\mathrm{(N)NLO}} = K^V_{\mathrm{(N)NLO}} - K^Z_{\mathrm{(N)NLO}}$
  → effectively degrades precision of last calculated order





Uncertainty estimates at NNLO QCD

Pure EW uncertainties



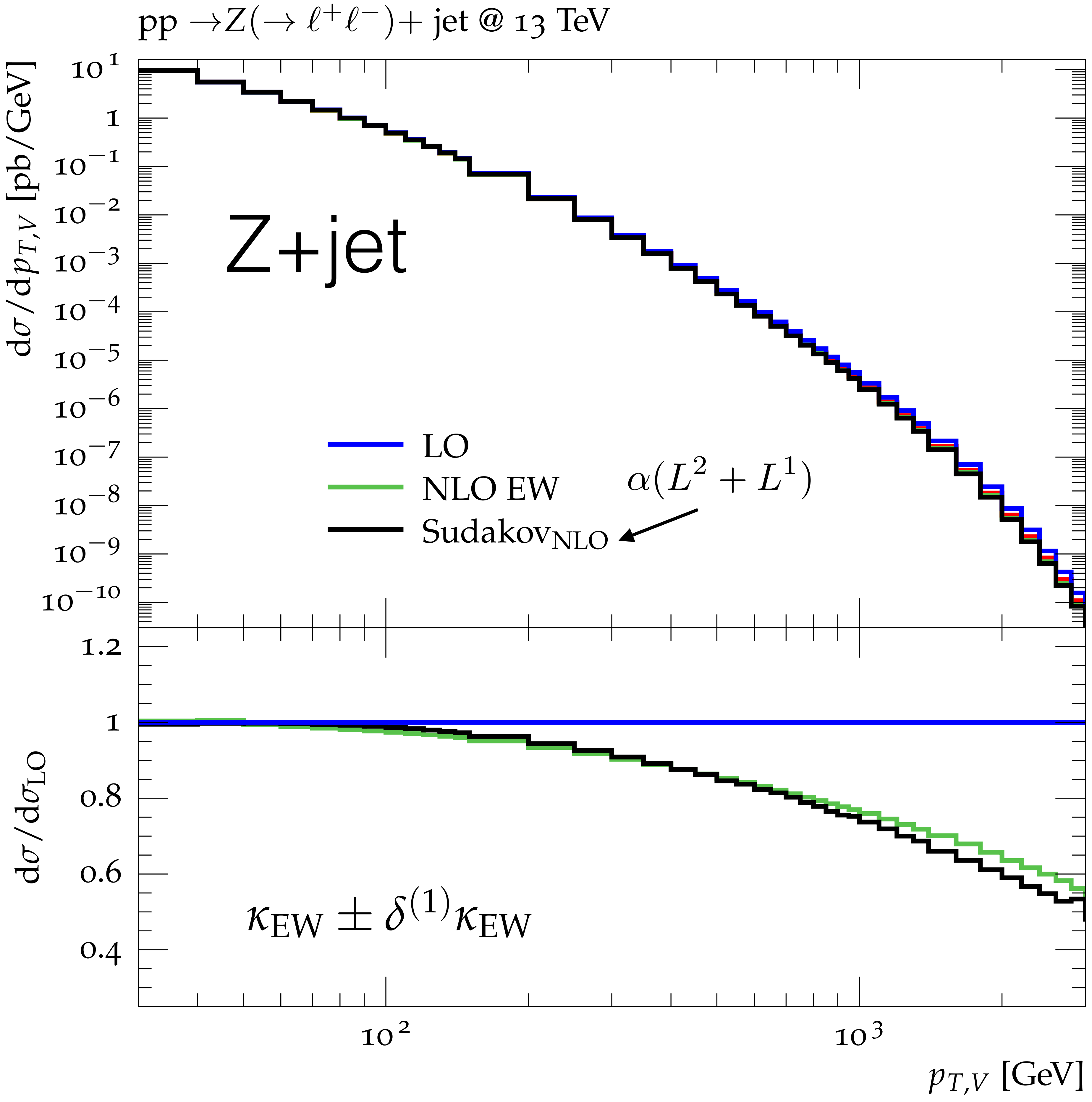
EW corrections become sizeable at large $p_{T,V}$: -30% @ 1 TeV

Origin: virtual EW Sudakov logarithms

How to estimate corresponding pure EW uncertainties of relative $\mathcal{O}(\alpha^2)$?

Pure EW uncertainties

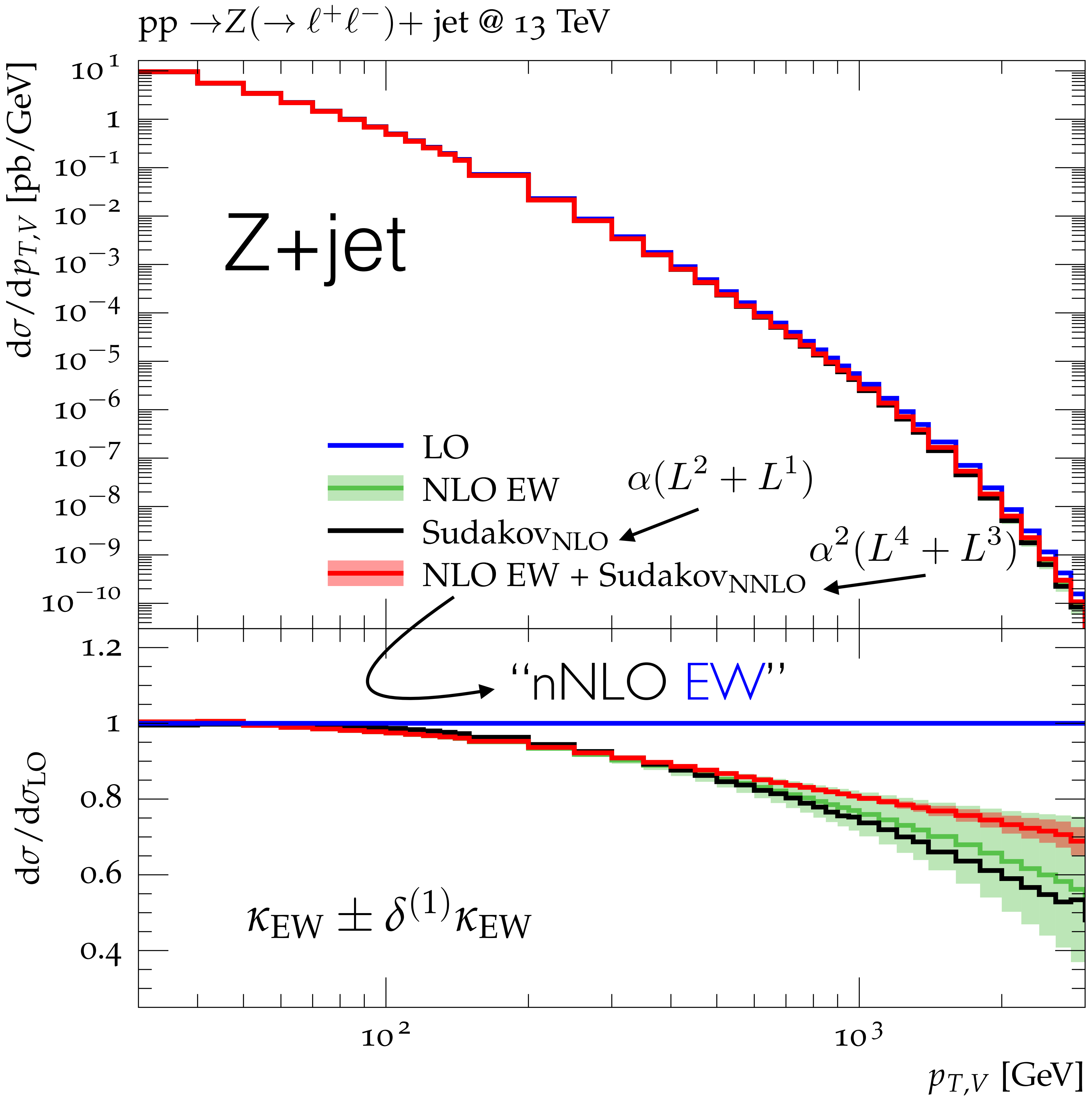
[JML et. al.: 1705.04664]



Large EW corrections dominated by Sudakov logs

Pure EW uncertainties

[JML et. al.: 1705.04664]



$$\kappa_{\mathrm{NLO\,EW}}(\hat{s},\hat{t}) = \frac{\alpha}{\pi}\left[\delta^{(1)}_{\mathrm{hard}} + \delta^{(1)}_{\mathrm{Sud}}\right]$$

$$\kappa_{\mathrm{NNLO\,Sud}}(\hat{s},\hat{t}) = \left(\frac{\alpha}{\pi}\right)^2 \delta^{(2)}_{\mathrm{Sud}}$$

Large EW corrections dominated by Sudakov logs

↓

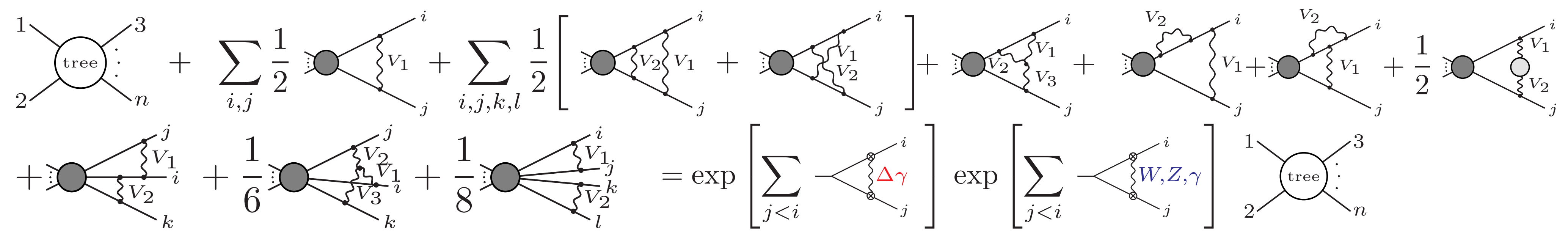
Uncertainty estimate of (N)NLO EW from naive exponentiation x 2:

$$\delta^{(1)}\kappa_{\mathrm{EW}} \simeq \frac{2}{k!}\left(\kappa_{\mathrm{NLO,EW}}\right)^k \qquad \textit{(correlated)}$$

↓

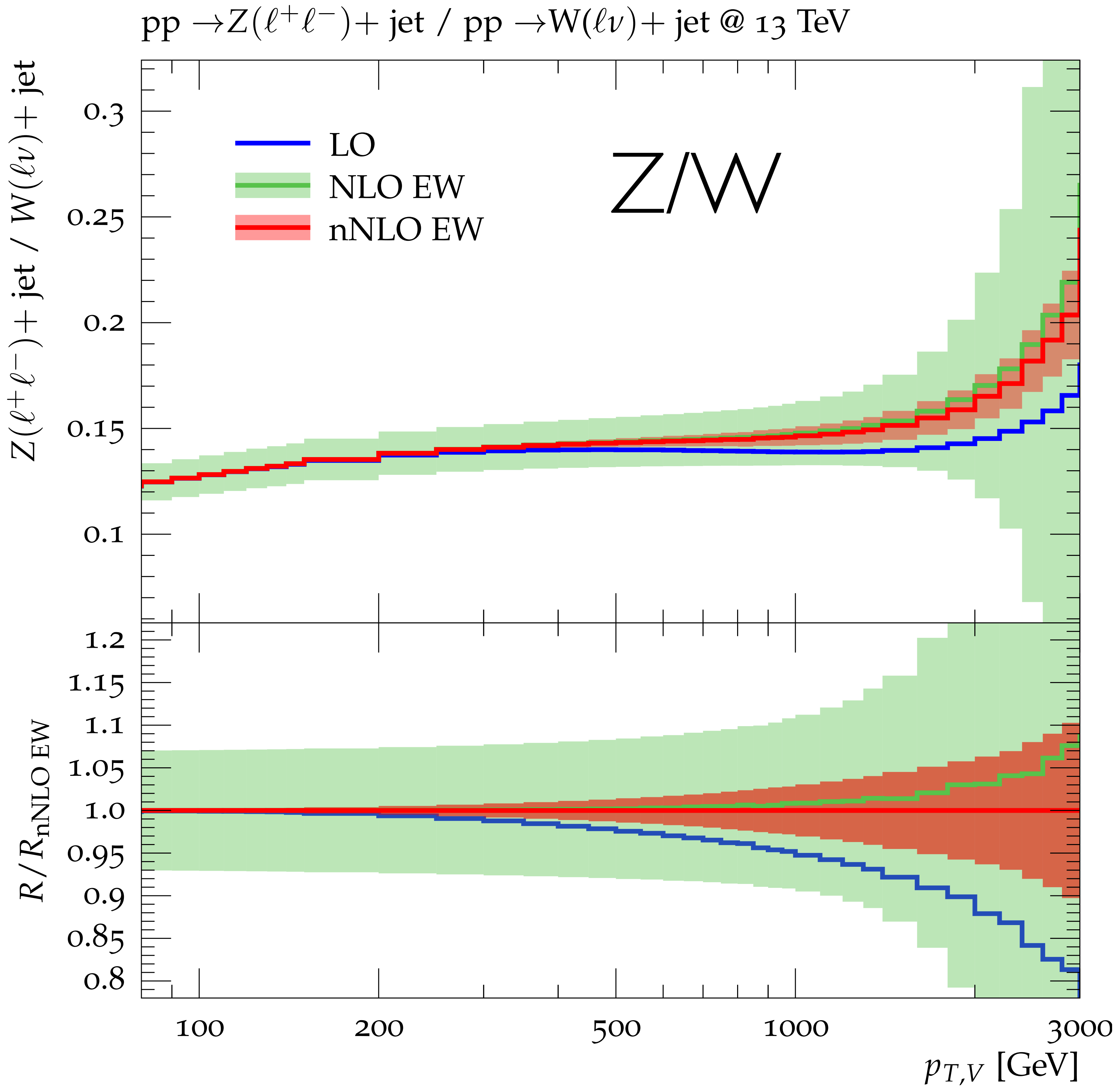
check against two-loop Sudakov logs

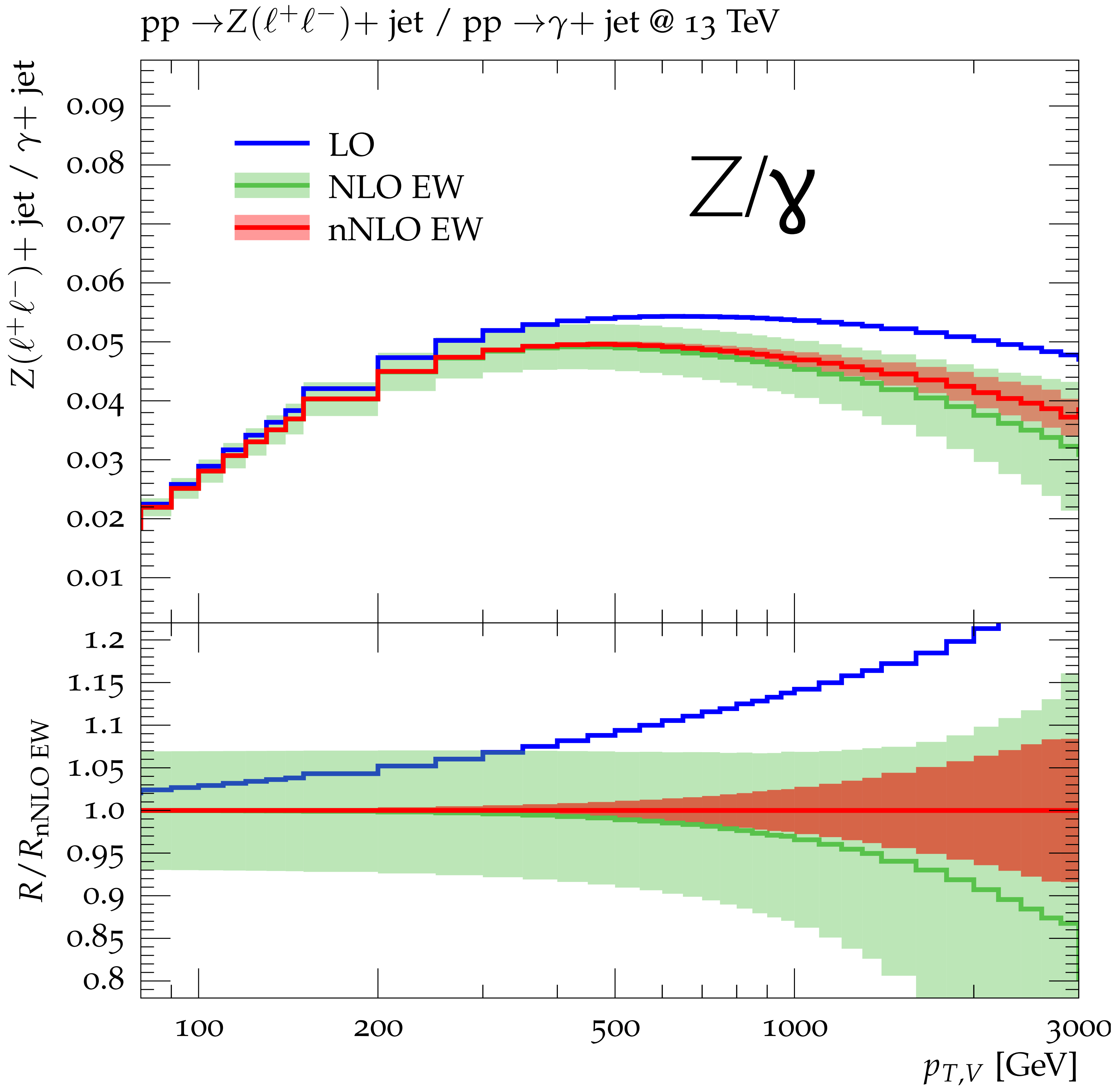
[Kühn, Kulesza, Pozzorini, Schulze; 05-07]



+ additional uncertainties for hard non-log NNLO EW effects *(uncorrelated)*

Pure EW uncertainties: ratios



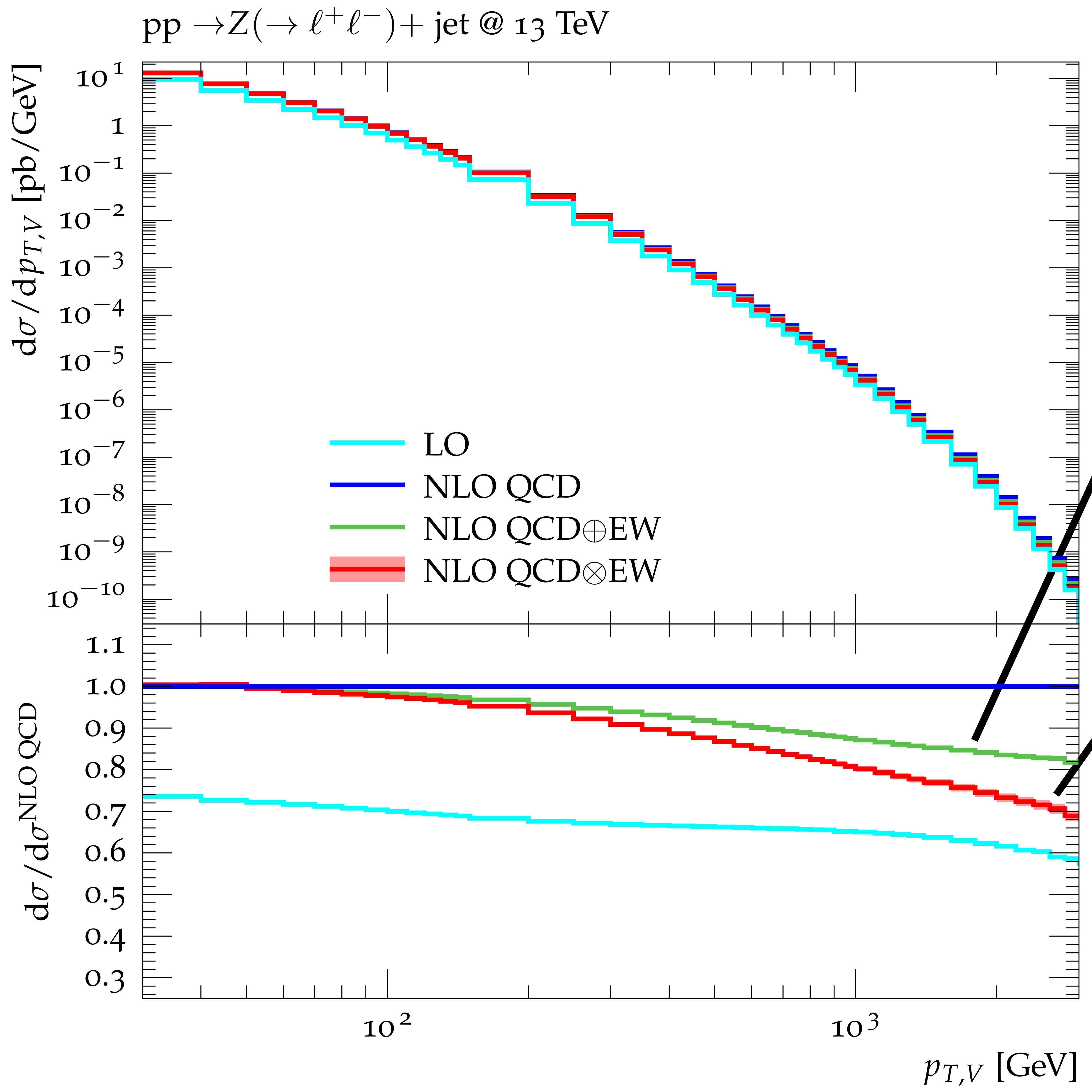


- NLO EW: ~5% for pT=1 TeV
- nNLO EW: ~1% for pT=1 TeV

- NLO EW: ~15% for pT=1 TeV
- nNLO EW: ~4% for pT=1TeV

δ(R) < 3-5% for pT < 1-2 TeV

Mixed QCD-EW uncertainties



Given QCD and EW corrections are sizeable, also mixed QCD-EW uncertainties of relative $\mathcal{O}(\alpha\alpha_s)$ have to be considered.

Additive combination

$$\sigma^{\rm NLO}_{\rm QCD+EW} = \sigma^{\rm LO} + \delta\sigma^{\rm NLO}_{\rm QCD} + \delta\sigma^{\rm NLO}_{\rm EW}$$

(no $\mathcal{O}(\alpha\alpha_s)$ contributions)

Multiplicative combination

$$\sigma^{\rm NLO}_{\rm QCD\times EW} = \sigma^{\rm NLO}_{\rm QCD}\left(1 + \frac{\delta\sigma^{\rm NLO}_{\rm EW}}{\sigma^{\rm LO}}\right)$$

(try to capture some $\mathcal{O}(\alpha\alpha_s)$ contributions, e.g. EW Sudakov logs × soft QCD)

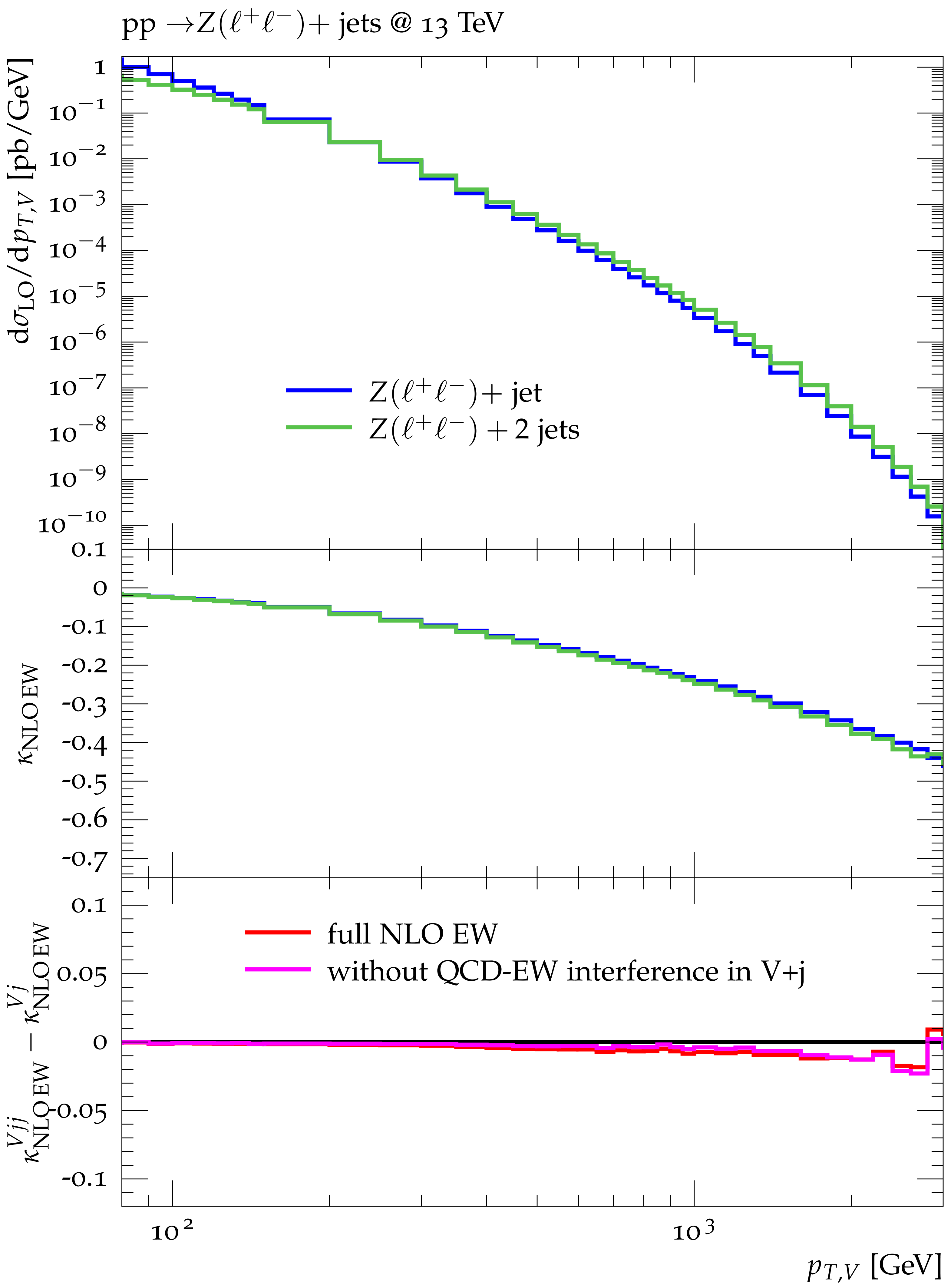
Difference between these two approaches indicates size of missing mixed EW-QCD corrections.

$$K_{\rm QCD\otimes EW} - K_{\rm QCD\oplus EW} \sim 10\% \quad \text{at 1 TeV}$$

Too conservative!?

For dominant Sudakov EW logarithms factorization should be exact!

Mixed QCD-EW uncertainties



Bold estimate:

Consider real $\mathcal{O}(\alpha\alpha_s)$ correction to V+jet

$\simeq$ **NLO EW to V+2jets**

and we observe

$$\frac{d\sigma_{\text{NLO EW}}}{d\sigma_{\text{LO}}}\bigg|_{V+2\text{jet}} - \frac{d\sigma_{\text{NLO EW}}}{d\sigma_{\text{LO}}}\bigg|_{V+1\text{jet}} \lesssim 1\%$$

strong support for

- factorization
- multiplicative QCD x EW combination

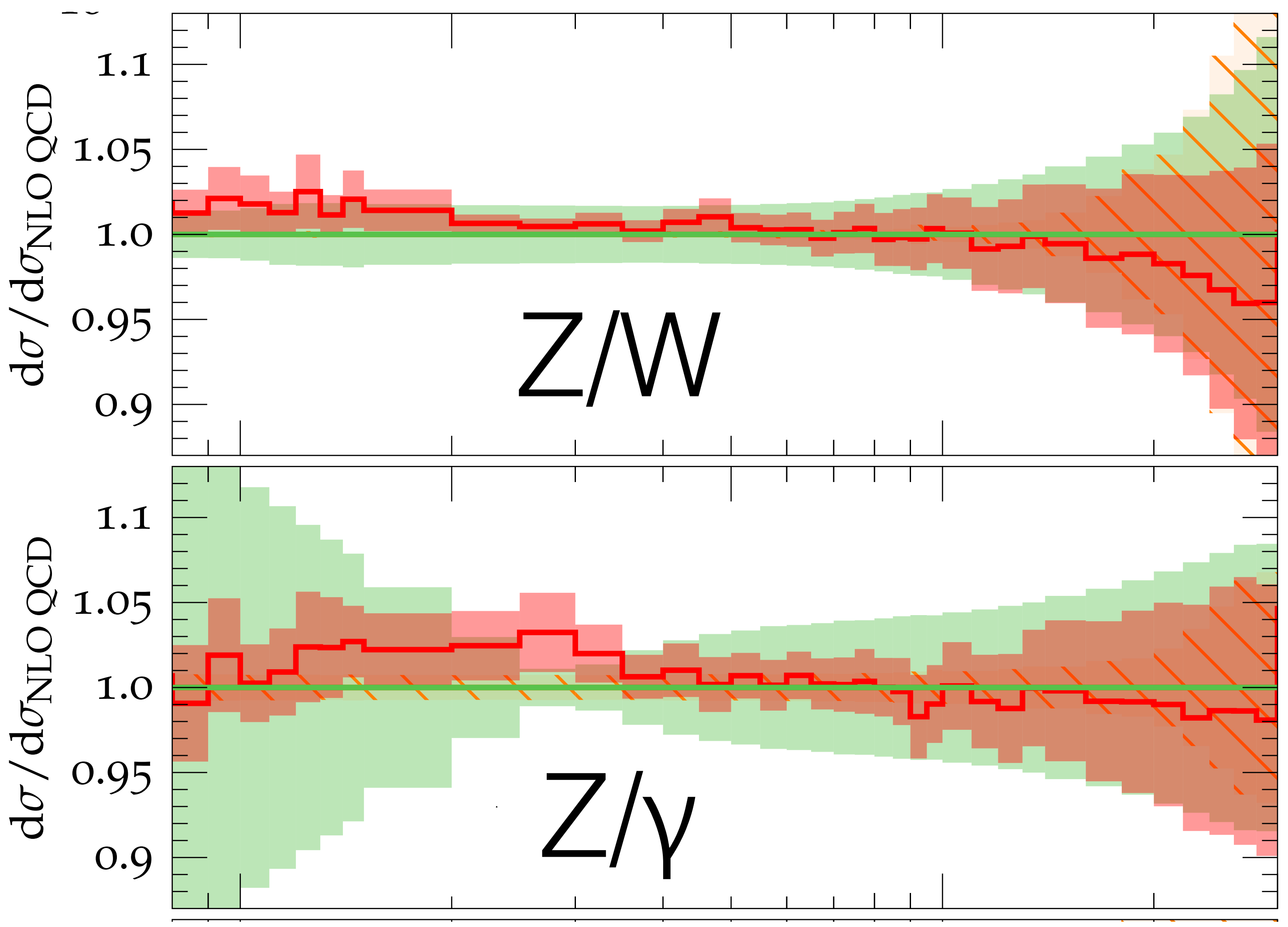
Estimate of non-factorising contributions

(correlated)

$$\delta K^{(V)}_{\text{mix}}(x) \quad = \quad 0.1\left[K^{(V)}_{\text{TH},\oplus}(x,\vec{\mu}_0) - K^{(V)}_{\text{TH},\otimes}(x,\vec{\mu}_0)\right]$$

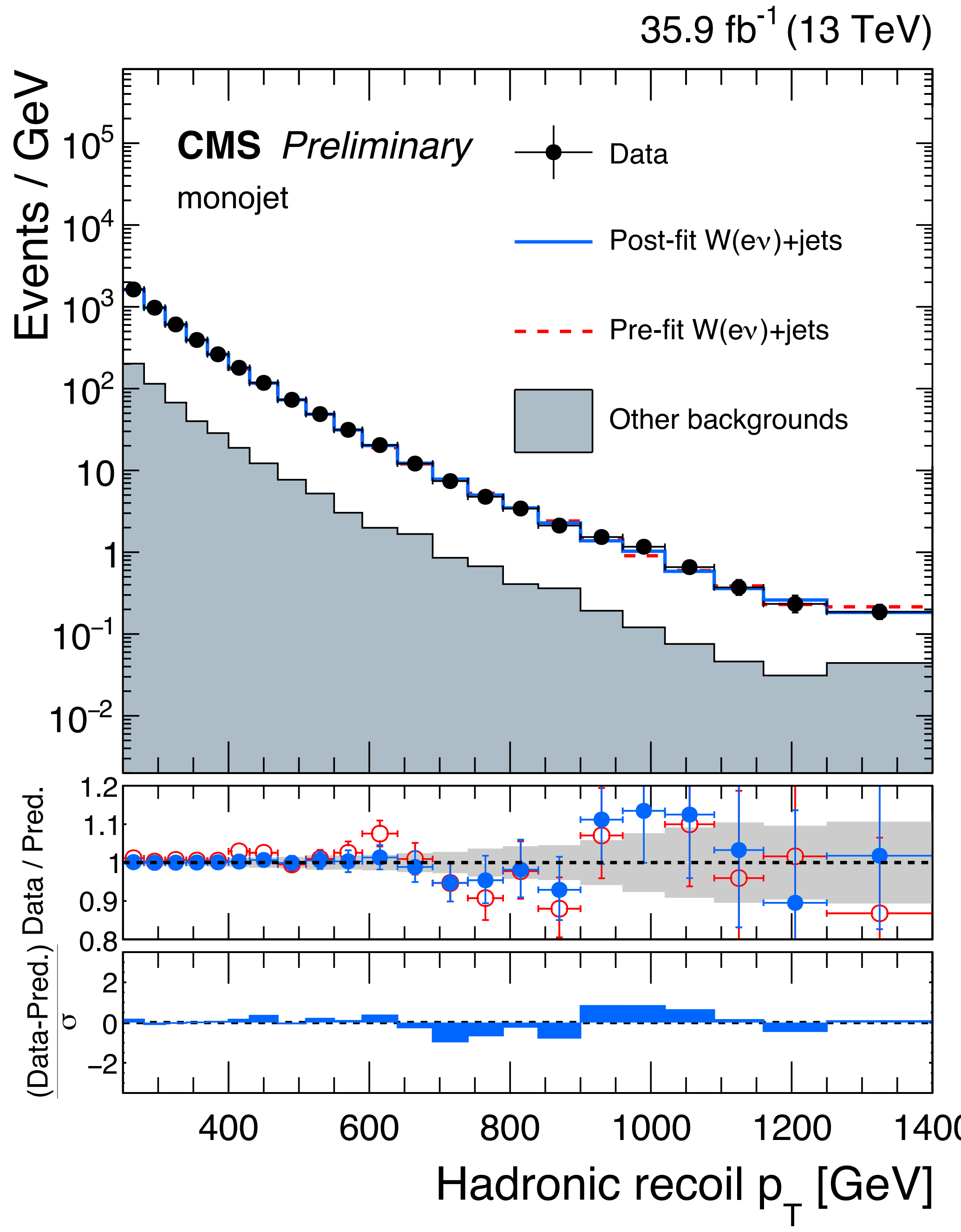
(tuned to cover above difference of EW K-factors)

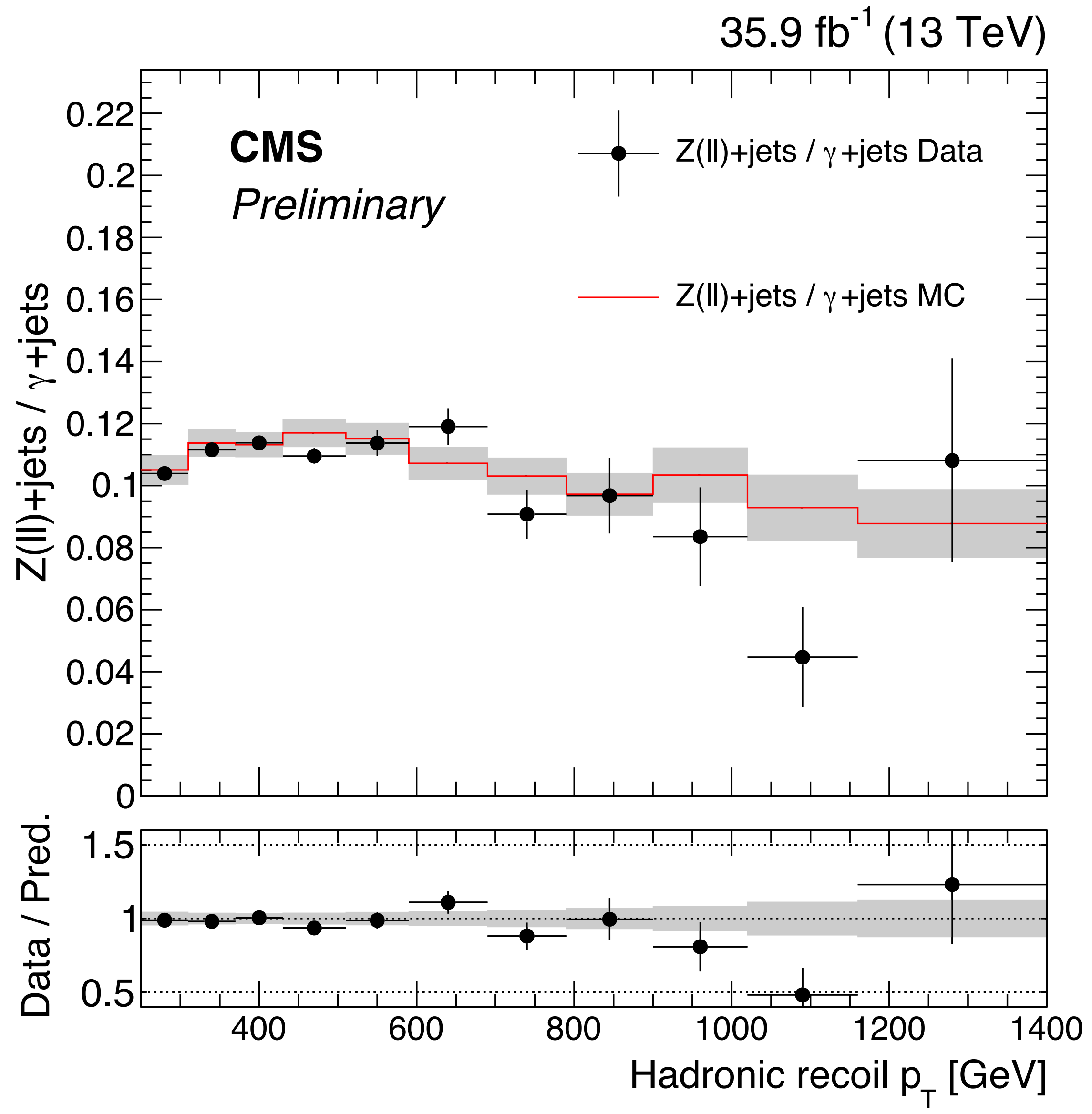
Combined uncertainties on V+jets ratios



- $\delta_{Z/W}$ =1-3% for pT < 1 TeV
- $\delta_{Z/\gamma}$ = 3-5% for pT < 1 TeV

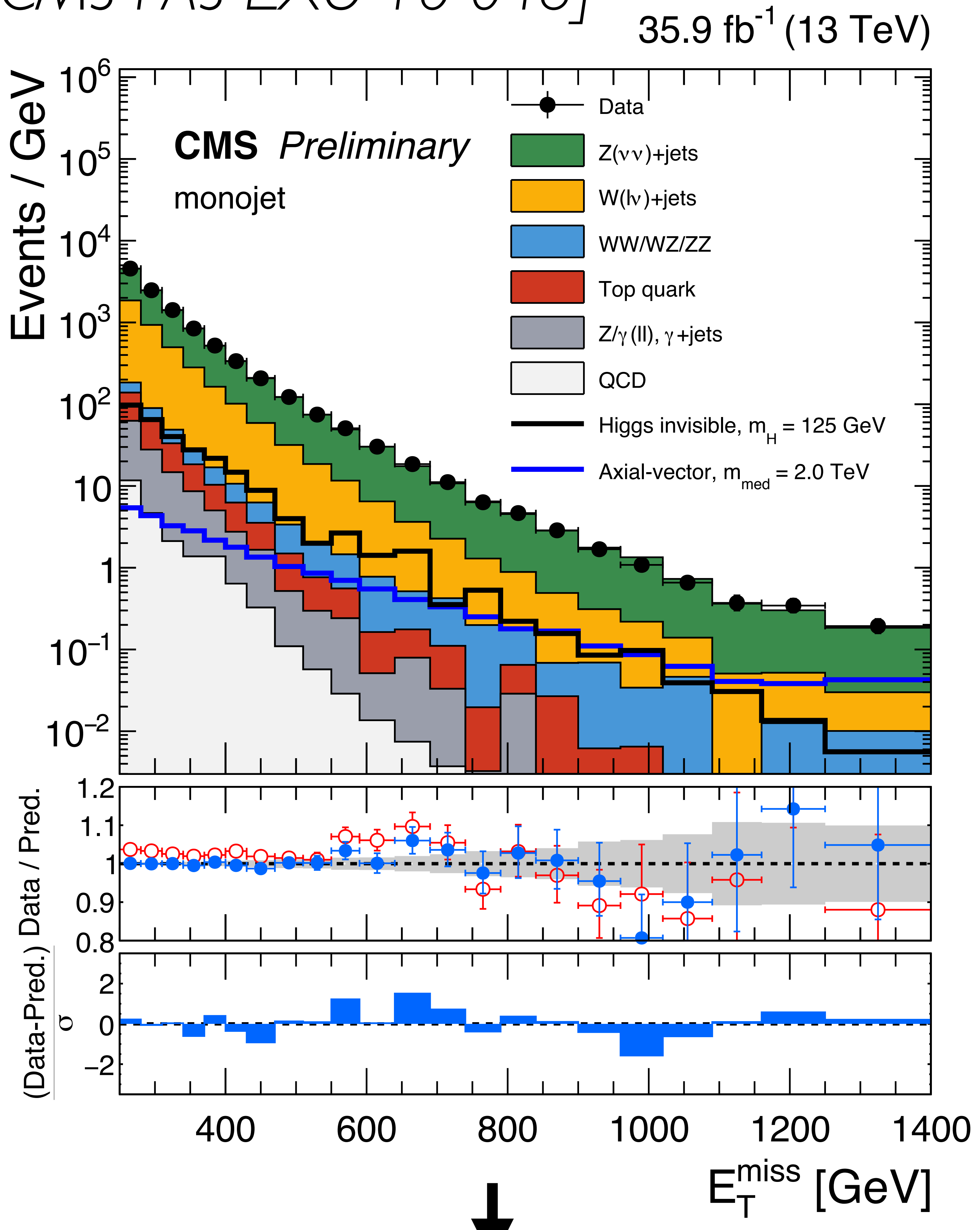
CR





SR

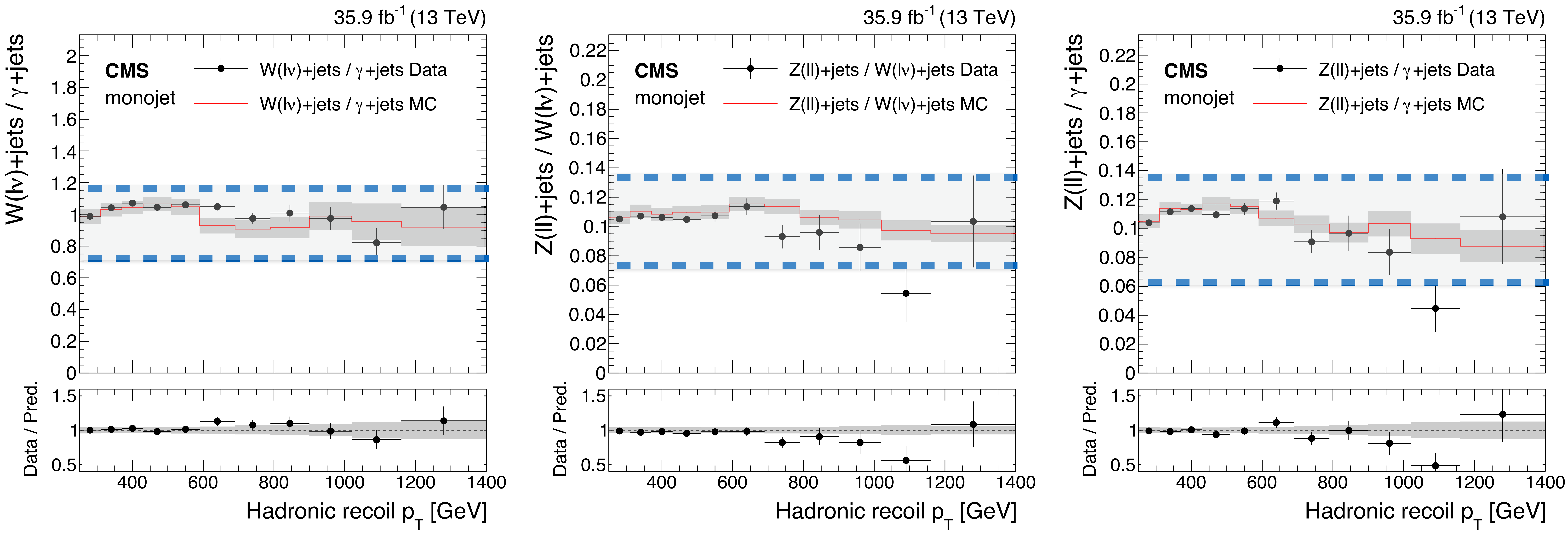
[CMS PAS EXO-16-048]



Unprecedented limits on monojet DM production!

Experimental closure tests
CMS monojet searches

[slide: Zeynep Demiragli, DM@LHC 2018]



Black ratio from data and statistical uncertainties / Red from MC

Grey band includes theoretical uncertainties

dashed lines -> what the uncertainties would have been without the work of the theory community

Conclusions

- There is no clear scale/signature for new physics effects: Let's explore the unknown leaving no stone unturned!
- Theory precision is often key to fully harness power of BSM searches.
- Detailed understanding of theory systematics is becoming pivotal.
- Automation of higher-order corrections allows for detailed phenomenological analyses for a multitude of process. But: need to look inside the black box.
- Let's push the precision frontier!

BACKUP

Alphabet-assisted bump hunt method

Applicable when both alphabet and bump hunt methods can be used

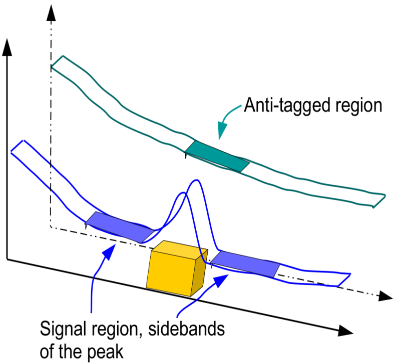
EXAMPLE: heavy resonances decaying into two Higgs bosons

Simultaneous fit of a control and signal region

→ Background normalization in signal region constrained by the pass/fail ratio from the alphabet method (sidebands region)

$$N_{\mathbf{SR}} = N_{\mathbf{CR}} \cdot R_{\mathbf{p/f}}$$

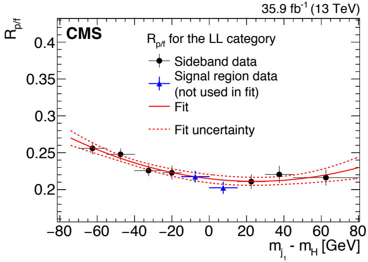
→ background shape extracted by fit using the same parametric function for both regions

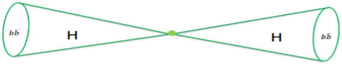


PRO: smooth fluctuations from the alphabet predictions

Alphabet-assisted bump hunt method

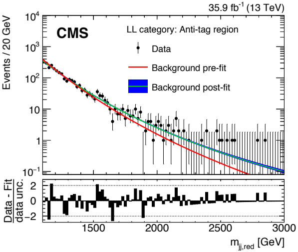
Applicable when both alphabet and bump hunt methods can be used

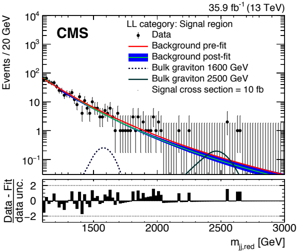




Selection of two large-cone double b-tagged jets

Dominant background: QCD multijet final states





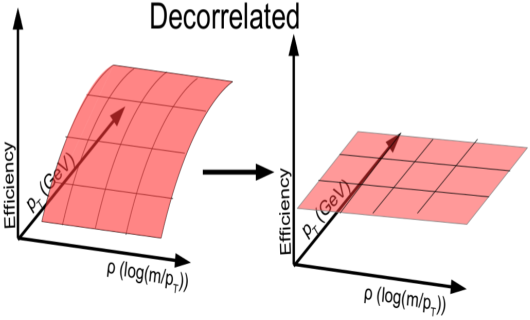
pre-fit: background fit in CR scaled by $R_{p/f}$ in SR

post-fit: simultaneous fit of CR and SR contrained by $R_{p/f}$

CMS - PLB 781 (2018) 244

Rhalphabet

2D fit of jet p_T and ρ (instead of mass), in order to avoid sculpting and correlation

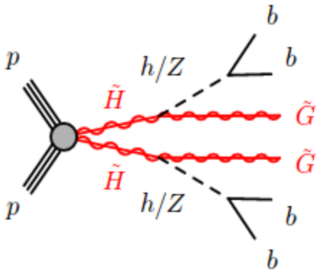


Decorrelation of variables avoid mass sculpting and achieves better performance (and smaller uncertainty)

Parametrize the pass/fail ratio in 2D (ρ and p_T)
(ideally it is flat, in reality has a little slope)

Factorization cuts: BDT-based event reweighting

Definition of a control region (two-tag) and extrapolation to four-tag region through a BDT (fully data-driven transfer function)



Search for higgsinos pair production with b-tagged jets

Low-mass analysis:

→ four b-tagged jets
→ reduced E_T^{miss}
→ jets paired according to their mass, being close to the Higgs mass

→ Main background from QCD multijet processes

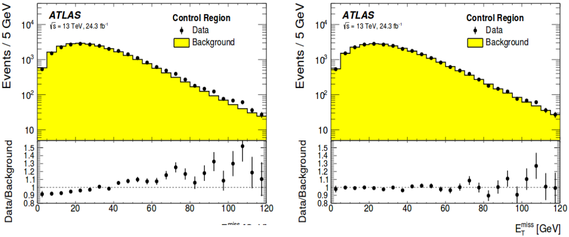
Control region defined by two b-tagged jets and two anti b-tag jets

Kinematic differences between 2-tag and 4-tag regions need to be corrected for → reweighting based on a BDT output

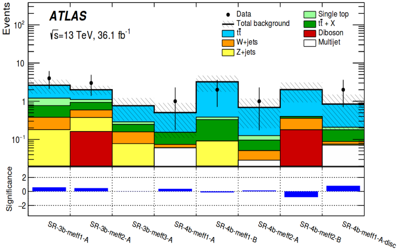
ATLAS - CERN-EP-2018-050 (Subm. to PRD)
CMS - PRD 97, 032007 (2018)

Factorization cuts: BDT-based event reweighting

Extrapolation to the signal region performed through a BDT regression based on 27 variables of Higgs candidates and event topology



Weights given by number of 2-tag and 4-tag events obtained for each endpoint BDT leaf



Reweighting crucial for proper background modelling

Three sources of background uncertainty:
- statistics of control regions
- closure of shape in control region
- closure of weights in validation regions

ATLAS - CERN-EP-2018-050 (Subm. to PRD)

Automation of NLO EW

MoCaNLO+Recola	$pp \to ll + 2$ jets $pp \to e^+e^-\mu^+\mu^- / \mu^+\mu^-\mu^+\mu^- / e^+\nu_e\mu^-\nu_\mu$ $pp \to e^+\nu_e\mu^-\nu_\mu\, bb\ (tt)$ $pp \to e^+\nu_e\mu^+\nu_\mu + 2$ jets (VBS) $pp \to e^+\nu_e\mu^-\nu_\mu\, bbH\ (t\bar{t}H)$	[1411.0916] [1601.07787] [1611.05338 [1607.05571] [1611.02951] [1708.00268] [1612.07138]
Sherpa/Munich+OpenLoops POWHEG+OpenLoops	$pp \to W+1,2,3$ jets $pp \to ll/l\nu/\nu\nu + 0, 1, 2$ jets (V+jets) $pp \to ll\nu\nu$ (VV) $pp \to llH/ll\nu H+0,1$ jet (HV)	[1412.5156] [1511.08692] [1705.00598] [1706.03522]
MadGraph_aMC@NLO +MadLoop	$pp \to t\bar{t}+H/Z/W$ $pp \to t\bar{t}$ $pp \to 2$ jets	[1504.03446] [1606.01915] [1705.04105] [1612.06548]
MadDipole+GoSam Sherpa+GoSam	$pp \to W+2$ jets $pp \to \gamma\gamma+0,1,2$ jets	[1507.08579] [1706.09022]

- many NLO QCD+EW calculations for **multi-particle processes** are becoming available
- NLO QCD+EW matching and merging with parton showers is under way (approximations available)
- Given the achieved automation: **attention is shifting towards detailed phenomenological applications**

Caveat: γ+jet

Note: this modelling of process correlations assumes close similarity of QCD effects between different V+jets processes

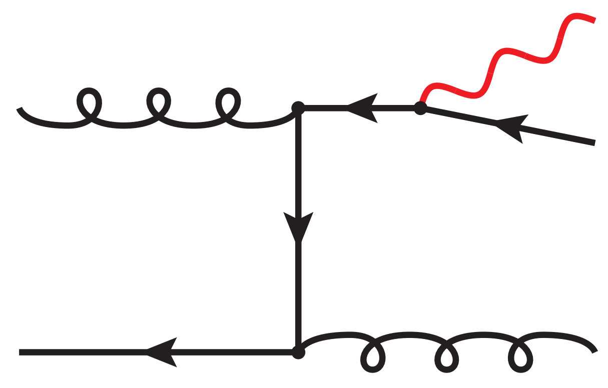
$$\left|\frac{\sigma^{(V)}_{\rm NLO}}{\sigma^{(V)}_{\rm LO}} - \frac{\sigma^{(Z)}_{\rm NLO}}{\sigma^{(Z)}_{\rm LO}}\right| \ll \left|\frac{\sigma^{(Z)}_{\rm NLO}}{\sigma^{(Z)}_{\rm LO}}\right|$$

- apart from PDF effects it is the case for W+jets vs. Z+jets
- at pT > 200 GeV it is in principle also the case for γ+jets vs. Z/W+jets

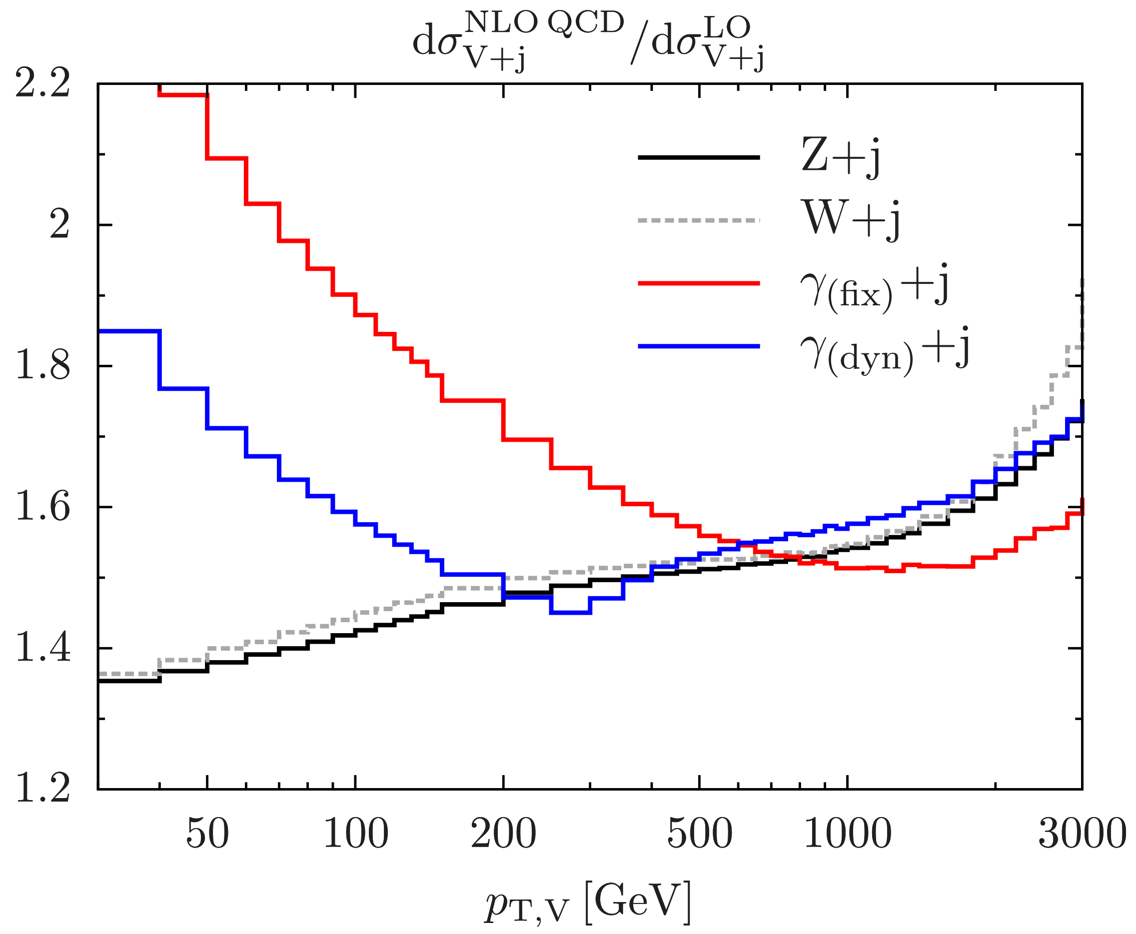
BUT: different logarithmic effects from fragmentation even at $pT \gg M_V$

W/Z+jet: mass cut-off → $\log(pT/M_V)$

γ+ jet: Frixione-isolation cone of radius R_0 → $\log(R_0)$



Consider dynamic γ-isolation with $R_{\rm dyn}(p_{{\rm T},\gamma}, \varepsilon_0) = \frac{M_Z}{p_{{\rm T},\gamma}\sqrt{\varepsilon_0}}$



- $\gamma_{\rm dyn}$ behaves like W or Z at $pT > M_V$

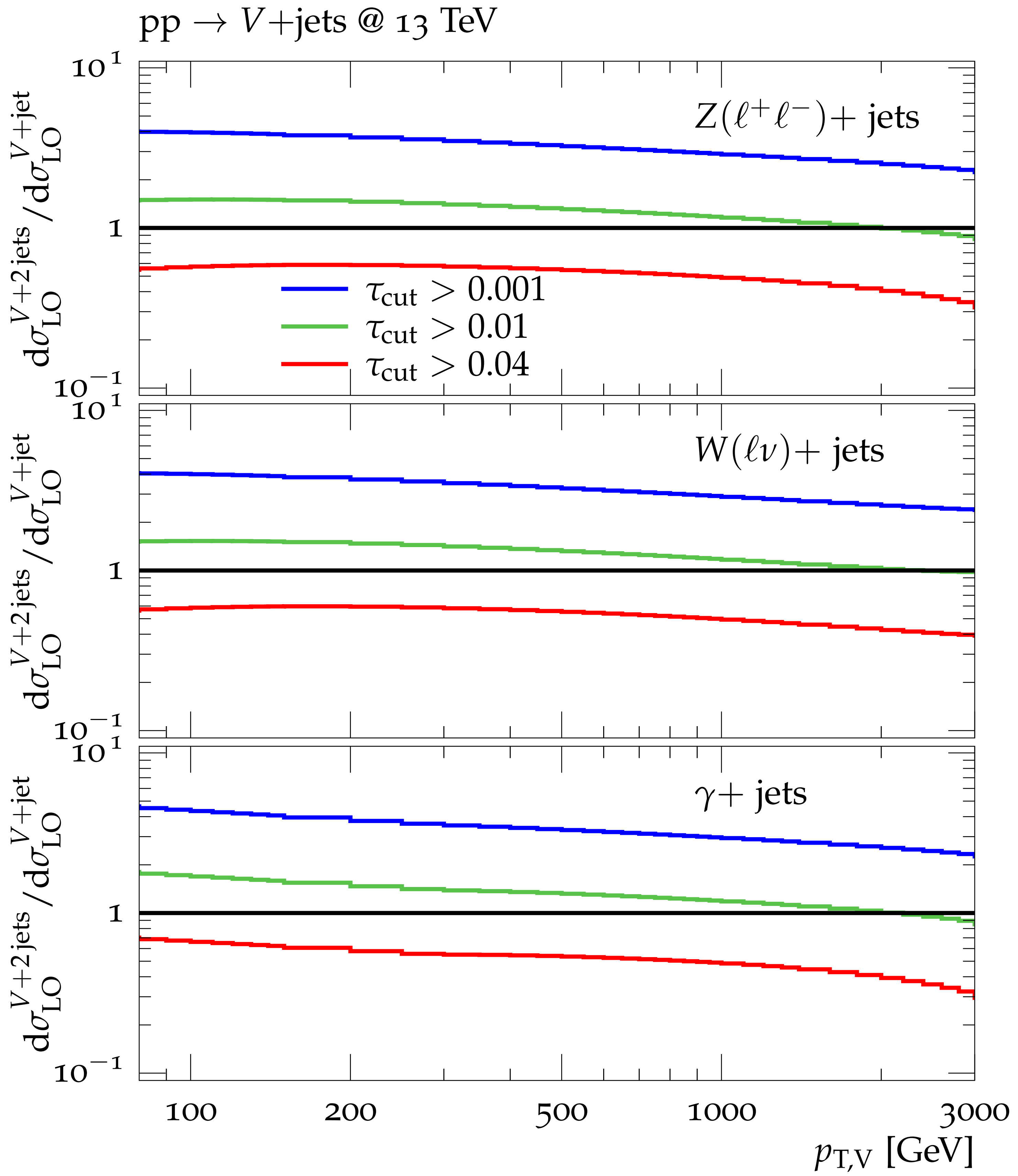
  ⇒justifies process-correlation estimate

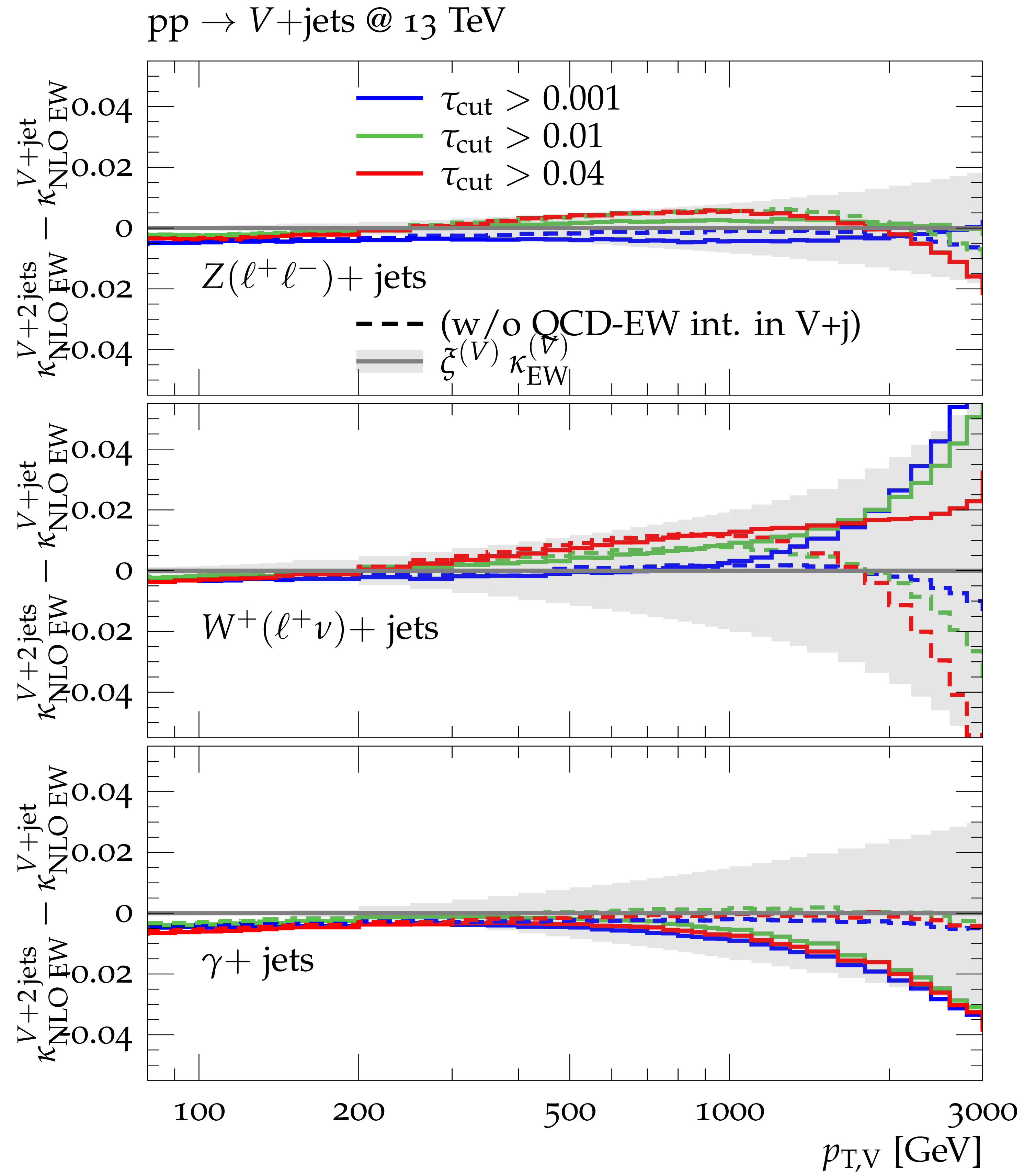
- Additional uncertainty: remnant part $\gamma_{\rm fix} - \gamma_{\rm dyn}$ (through extra MC reweighting)

(uncorrelated)

Mixed QCD-EW uncertainties

Estimate of non-factorising contributions





N-jettiness cut ensures approx. constant ratio V+2jets/V+jet

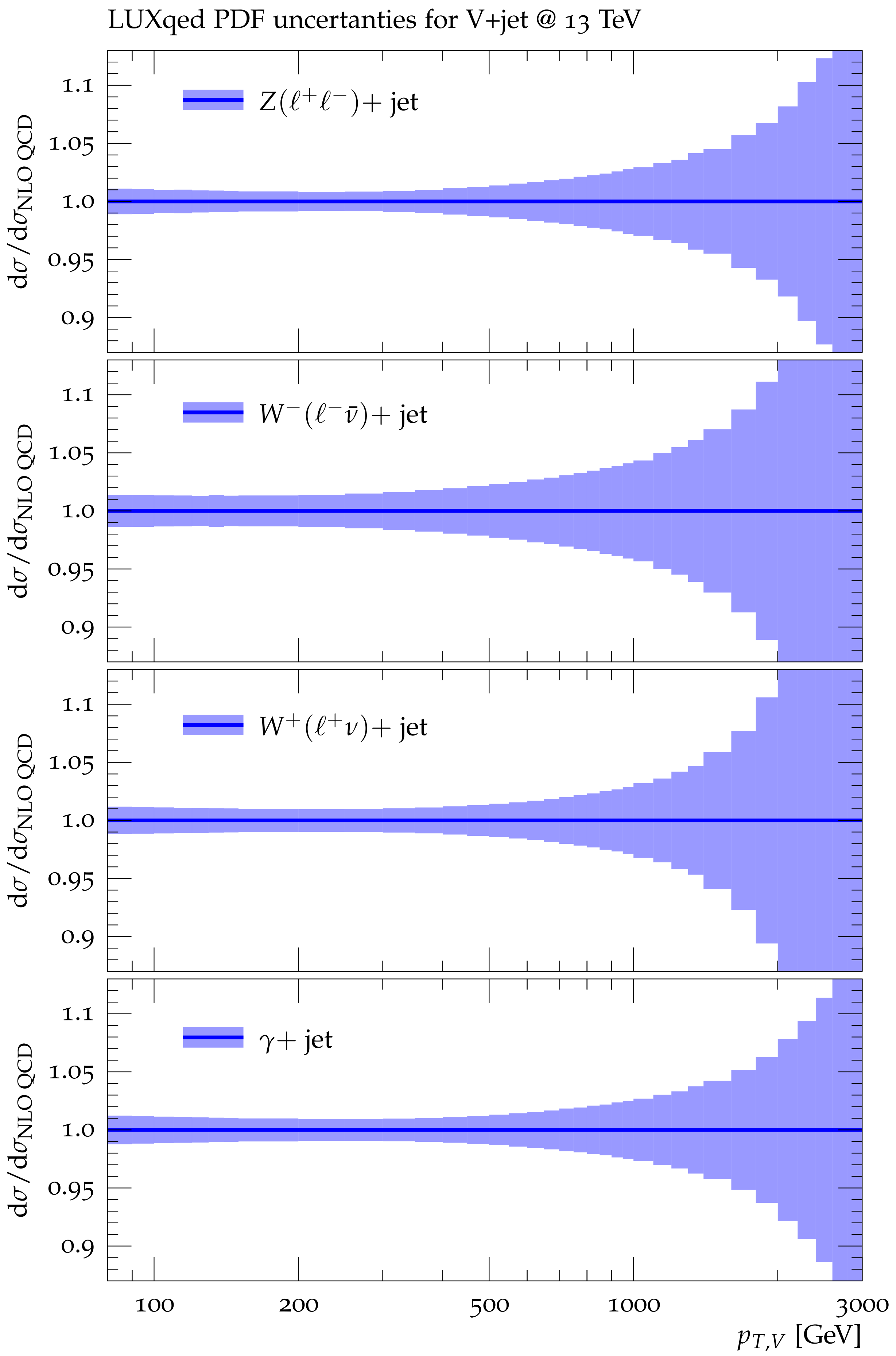
$$\tau_1 = \sum_k \min_i \left\{ \frac{2p_i \cdot q_k}{Q_i \sqrt{\hat{s}}} \right\}$$

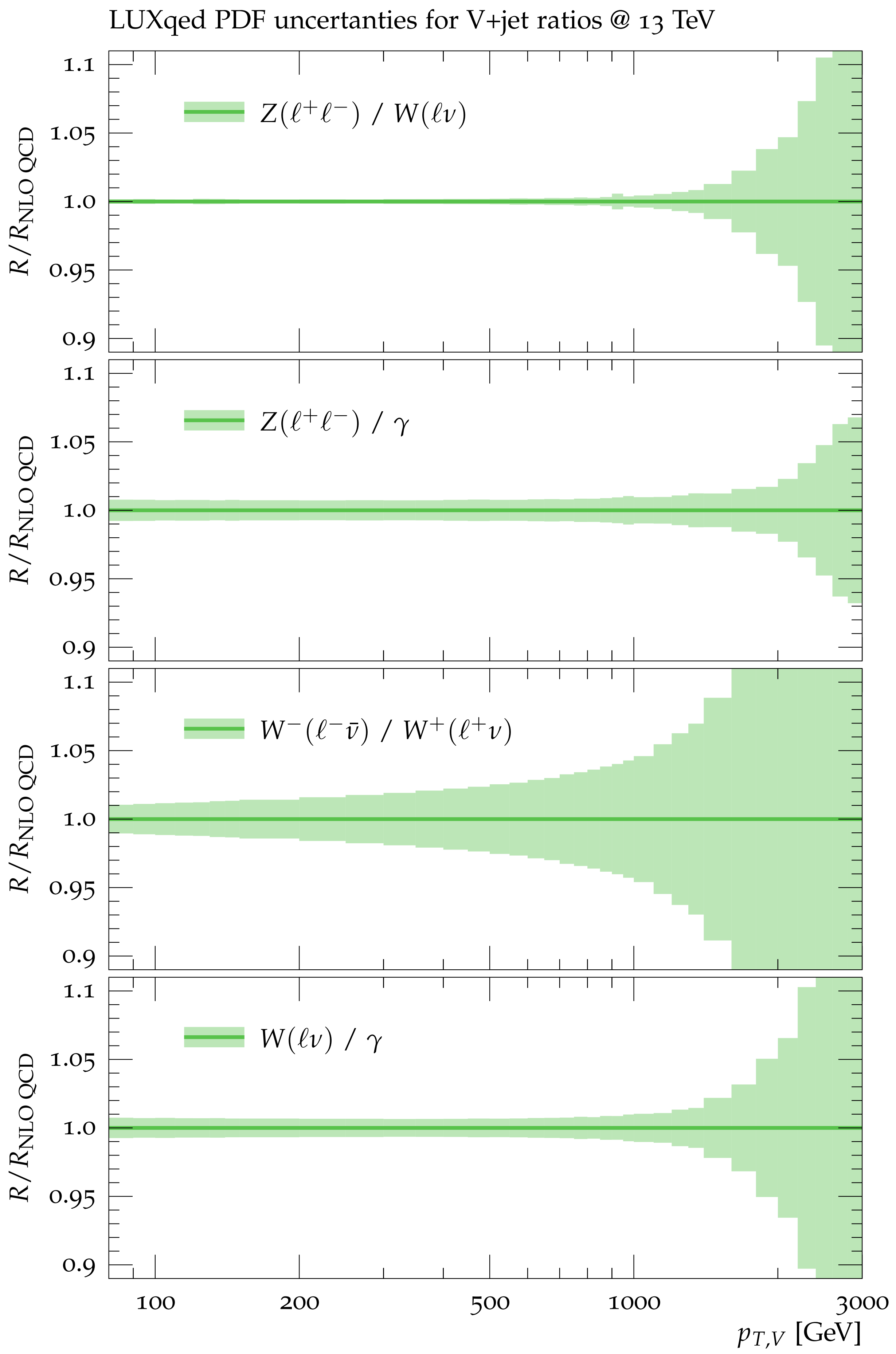
$$\delta K^{(V)}_{\text{mix}}(x, \boldsymbol{\mu}) = \xi^{(V)} \left[K^{(V)}_{\text{TH},\otimes}(x, \boldsymbol{\mu}) - K^{(V)}_{\text{TH},\oplus}(x, \boldsymbol{\mu}) \right]$$

$$\xi^Z = 0.1, \qquad \xi^W = 0.2, \qquad \xi^\gamma = 0.4$$

(tuned to cover above difference of EW K-factors)

PDF uncertainties (LUXqed=PDF4LHC)





- δ_{PDF} < 2 % for pT,V < 800 GeV
- δ_{PDF} < 5 % for pT,V < 1500 GeV

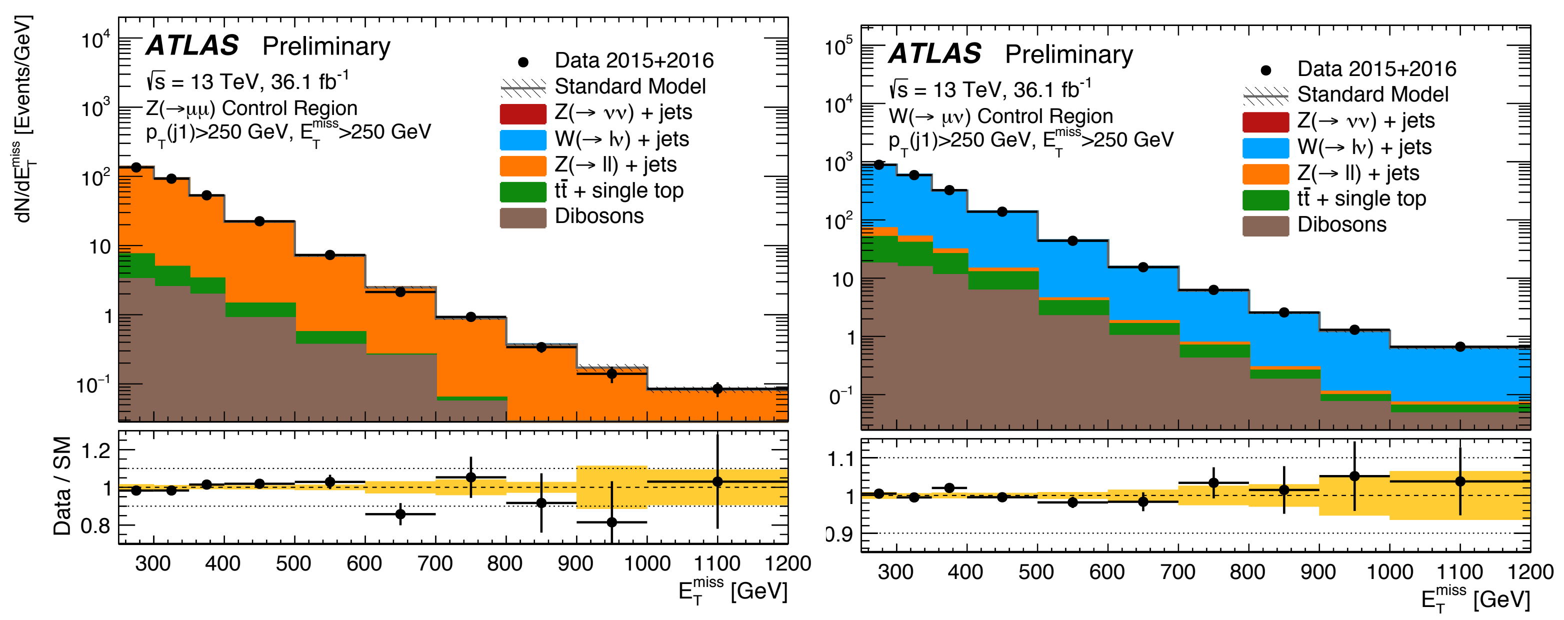
- Z/W: δ_{PDF} < 0.5%(2%) for pT,V < 800 GeV(1.5 TeV)
- Z/**γ** & W/**γ**: δ_{PDF} < 2% for pT,V < 1.3 TeV
- W^-/W^+: δ_{PDF} > 5% for pT,V < 1 TeV
  (due to large uncertainties on u/d ratio at large Bjorken-x)

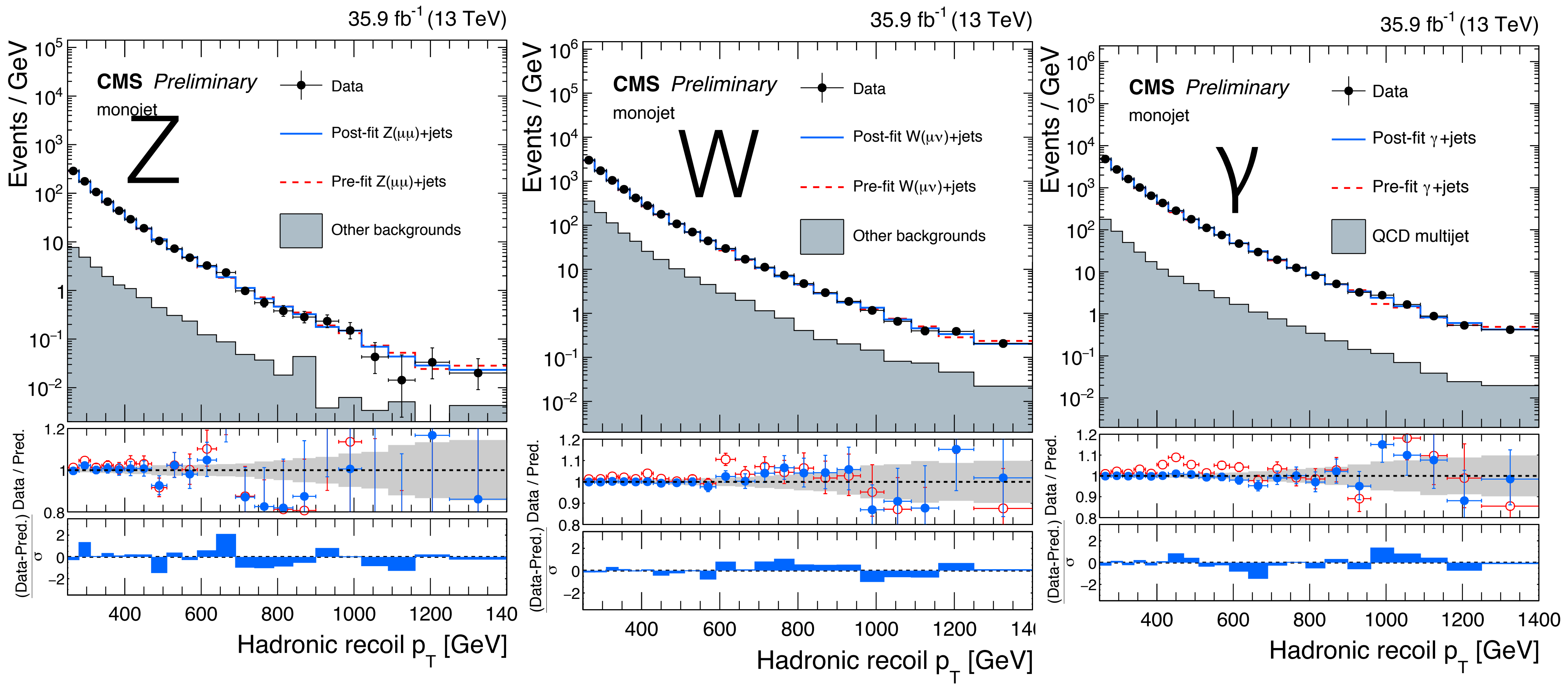
Experimental closure tests
in recent ATLAS & CMS monojet searches

[ATLAS-CONF-2017-060]

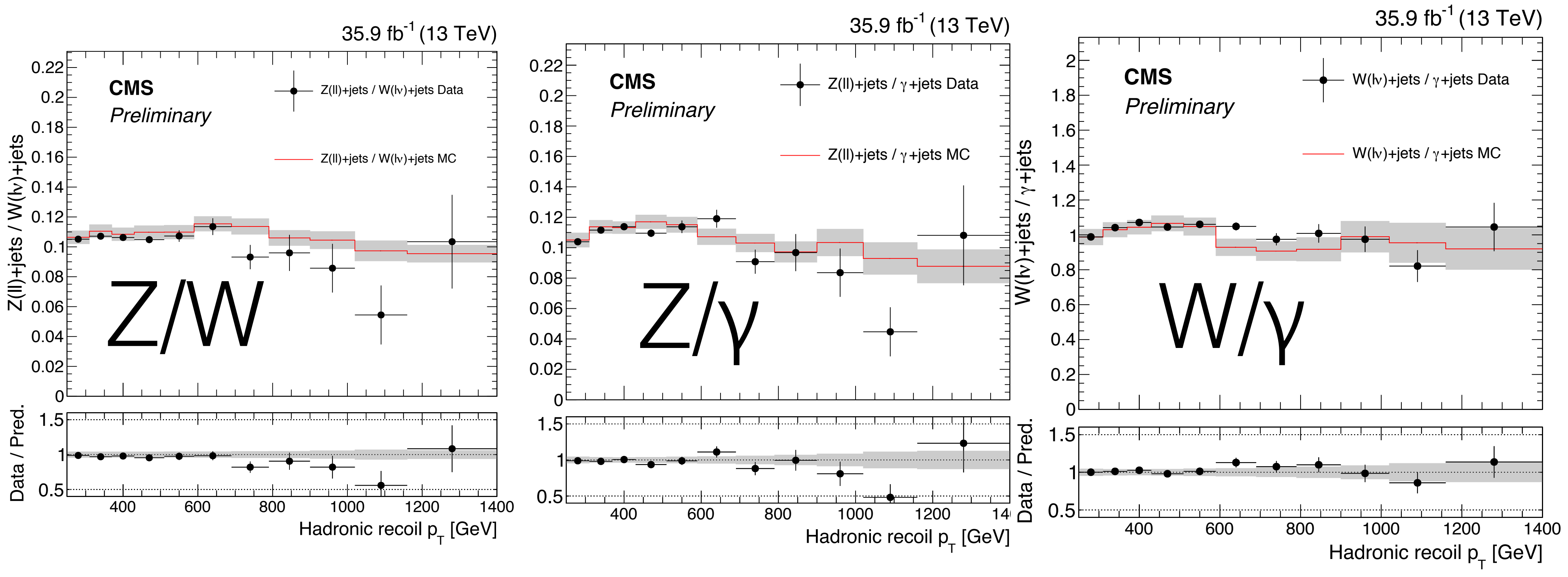
[CMS PAS EXO-16-048]

Nominals

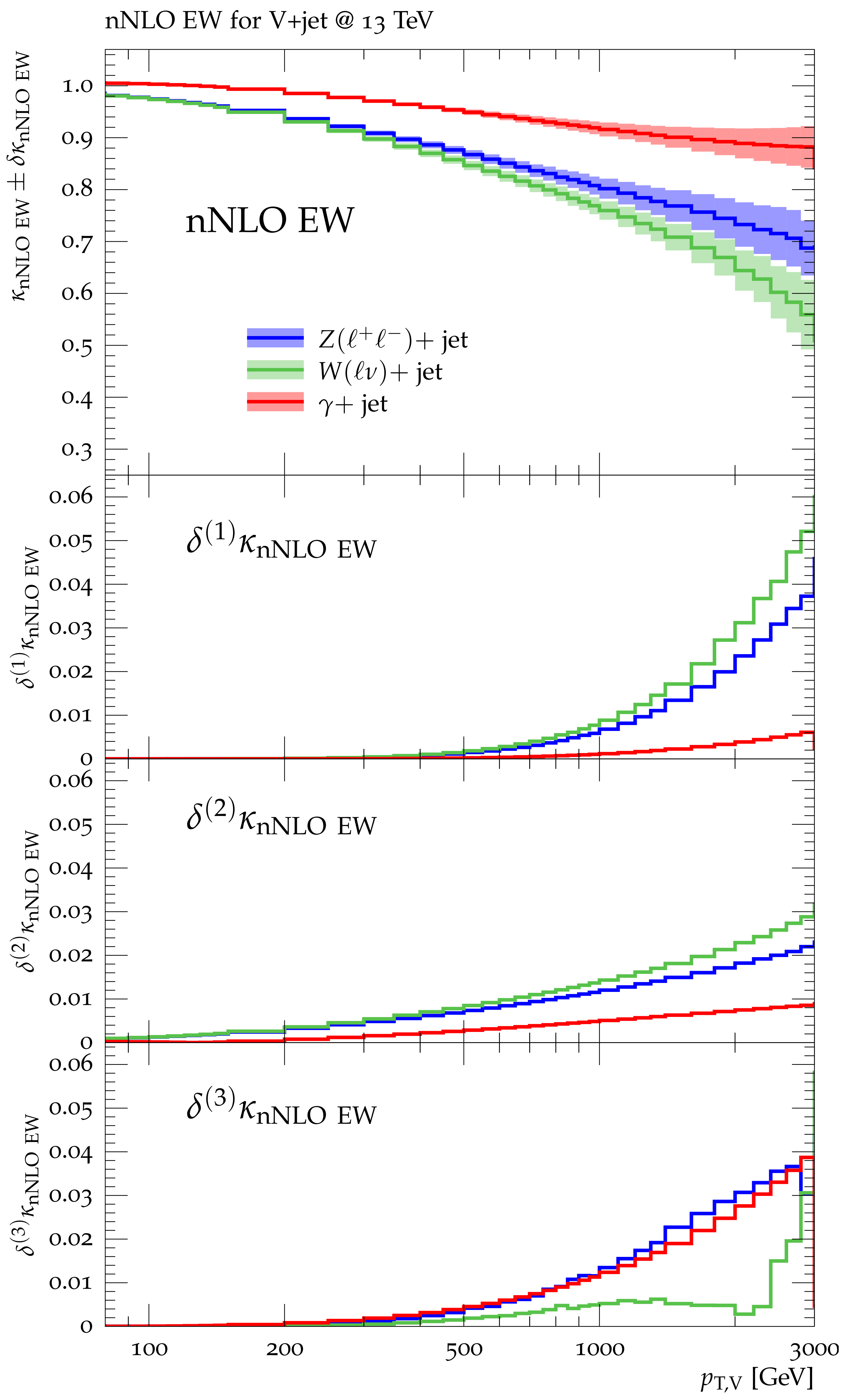




Ratios



Pure EW uncertainties



nNLO EW corrections at 1 TeV

- -10% for γ+jets
- -20% for Z+jet
- -25% for W+jet

$$d\sigma_{\mathrm{EW}} = \left[1+\frac{\alpha}{\pi}\left(\delta^{(1)}_{\mathrm{hard}}+\delta^{(1)}_{\mathrm{Sud}}\right)+\left(\frac{\alpha}{\pi}\right)^2\left(\delta^{(2)}_{\mathrm{Sud}}+\delta^{(1)}_{\mathrm{Sud}}\delta^{(1)}_{\mathrm{hard}}+\delta^{(1)}_{\mathrm{hard}}\delta^{(1)}_{\mathrm{hard}}+\delta^{(2)}_{\mathrm{hard}}\right)+\left(\frac{\alpha}{\pi}\right)^3\left(\delta^{(3)}_{\mathrm{Sud}}\ldots\right)\right]$$

- ‘higher-order Sudakov logs’

$$\delta^{(1)}\kappa^{(V)}_{\mathrm{EW}}(x) \quad =: \frac{2}{3}\kappa^{(V)}_{\mathrm{NLO\,EW}}(x)\,\kappa^{(V)}_{\mathrm{NNLO\,Sud}}(x)$$

(correlated)

Additional uncorrelated uncertainties:

- ‘hard non-log NNLO EW effects I’

(uncorrelated)

$$\delta^{(2)}\kappa^{(V)}_{\mathrm{EW}}(x) = 0.05\,\kappa^{(V)}_{\mathrm{NLO\,EW}}(x)$$

$$\Leftrightarrow \quad \delta^{(2)}_{\mathrm{hard}} \leq \frac{0.05\pi}{\alpha}\delta^{(1)}_{\mathrm{hard}} \simeq 20\,\delta^{(1)}_{\mathrm{hard}}$$

- ‘hard non-log NNLO EW effects II’

$$\delta^{(3)}\kappa^{(V)}_{\mathrm{EW}}(x) \quad = \quad \kappa^{(V)}_{\mathrm{NNLO\,Sud}}(x) - \frac{1}{2}[\kappa^{(V)}_{\mathrm{NLO\,EW}}(x)]^2$$

(uncorrelated)

estimate of typical size of $\left[\delta^{(1)}_{\mathrm{hard}}\right]^2$ or $\delta^{(1)}_{\mathrm{hard}} \times \delta^{(1)}_{\mathrm{Sud}}$.