# MEASUREMENT OF THE Z BOSON PRODUCTION IN ASSOCIATION WITH AT LEAST TWO B JETS IN PP COLLISIONS AT $\sqrt{s}$ = 13 TeV

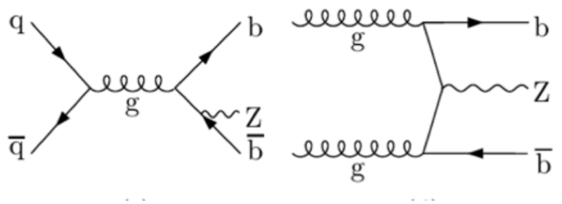

Hsin-Wei Hsia

for the CMS Collaboration

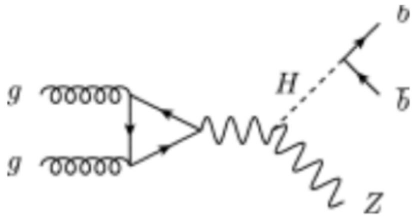
University at Buffalo The State University of New York

# Introduction

- The Z boson production associated with jets from b quarks provides important tests to perturbative quantum chromodynamics calculations.
- It is also the major background for many physics processes, including Higgs boson production in association with a Z, ZH(H $\rightarrow b\bar{b}$), and searches for new physics phenomena.




Examples of Feynman diagrams of Z+2b production




Example of Feynman diagram of gluon-initiated ZH production

Previous measurements:

Z+b jet cross section measurements at 7 TeV and 8 TeV by CMS Collaboration.

Ratios of cross sections, σ(Z+c jet)/σ(Z+jets), σ(Z+b jet)/σ(Z+jets) and σ(Z+c jet)/σ(Z+b jets) at 13 TeV by CMS Collaboration.

Cross section measurement of Z+b jet at 13 TeV (2016) by ATLAS Collaboration.

# Outline

- Overview of the measurements
- Background estimation for Z + $\geq$ 2 b jets
- Uncertainties in integrated cross section for Z + $\geq$ 2 b jets
- Integrated cross section for Z + $\geq$ 2 b jets
- Unfolded differential & normalized differential cross section distributions
- Summary

# Overview of the measurements

- Full Run 2 collision data with integrated luminosity of 137.1 $\mathrm{fb}^{-1}$ (2016, 2017, 2018).
  The cross section is measured in the leptonic decay channel of Z bosons (Z→ $ee$ or Z→ $\mu\mu$ ).

- Trigger: Single lepton trigger
  - at least one electron (muon) candidate with minimum $p_{\mathrm{T}}$ of 27, 32, and 32 GeV (24, 27, and 24 GeV) during 2016, 2017, and 2018 data-taking periods.

- Muon: isolated muon with leading (sub-leading) muon $p_{\mathrm{T}}$ > 35 GeV (25 GeV), $|\eta| < 2.4$
- Electron: isolated electron with leading (sub-leading) electron $p_{\mathrm{T}}$ > 35 GeV (25 GeV)
  $|\eta| < 2.4$
- Z boson: pair of oppositely charged leading and sub-leading leptons within the mass window
  $71\ \mathrm{GeV} < M_{ll} < 111\ \mathrm{GeV}$

- Jet: jets reconstructed with anti-$k_T$ algorithm with a distance parameter of 0.4
  $p_{\mathrm{T}}$ > 30 GeV, $|\eta|$ < 2.4, lepton overlap removal, tight jet ID
  jets from pileup events are excluded
- b jet: b jets tagged with DeepCSV tagger
  - tight tagging requirement is implemented to reduce the backgrounds (reduce DY+ bX, XX backgrounds, X = c or light jets)
    - loose, medium, tight operating point corresponds to approximately 80%, 60%, 50% b quark tagging efficiency, and 10%, 1%, 0.1% misidentification rate for light jets.
- MET < 50 GeV → to reduce $t\bar{t}$ background
- The observed distributions at reconstruction level are unfolded to particle level ones by applying corrections for the detector effects.

# Events in Z + ≥ 2 b jets final states

| | 2016 | | 2017 | | 2018 | |
|---|---|---|---|---|---|---|
| | Electron | Muon | Electron | Muon | Electron | Muon |
| DY+bb | 1976 | 4206 | 2241 | 4557 | 4361 | 8934 |
| DY+bX | 38 | 89 | 29 | 81 | 72 | 190 |
| DY+XX | 10 | 89 | 2 | 55 | 43 | 237 |
| tt | 784 | 1748 | 1102 | 2362 | 1825 | 3865 |
| single t | 18 | 40 | 22 | 59 | 48 | 99 |
| VV | 44 | 97 | 65 | 136 | 98 | 218 |
| ZH | 11 | 22 | 13 | 27 | 23 | 45 |
| MC | 2880 | 6291 | 3473 | 7278 | 6470 | 13589 |
| Data | 2549 | 5503 | 3070 | 6554 | 5070 | 10757 |
| Data/MC | 0.89 | 0.87 | 0.88 | 0.90 | 0.78 | 0.79 |

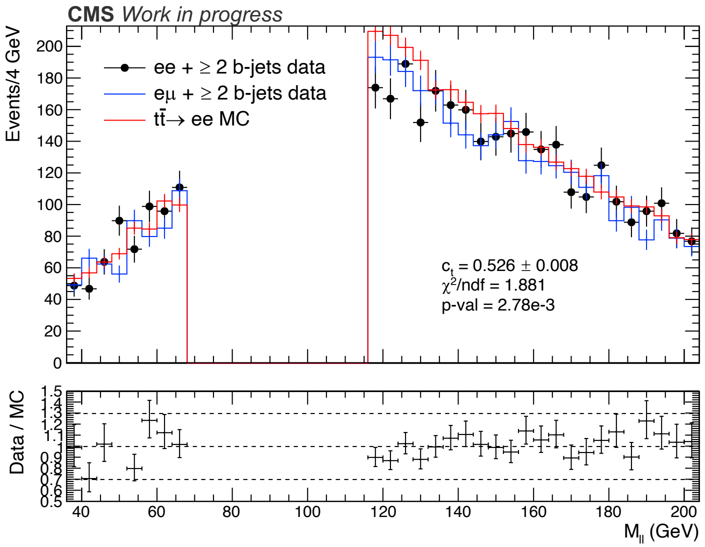
- Signal (DY+bb) is about 66% of total selected events.
- $t\bar{t}$ is the significant background (~30% of total selected events).

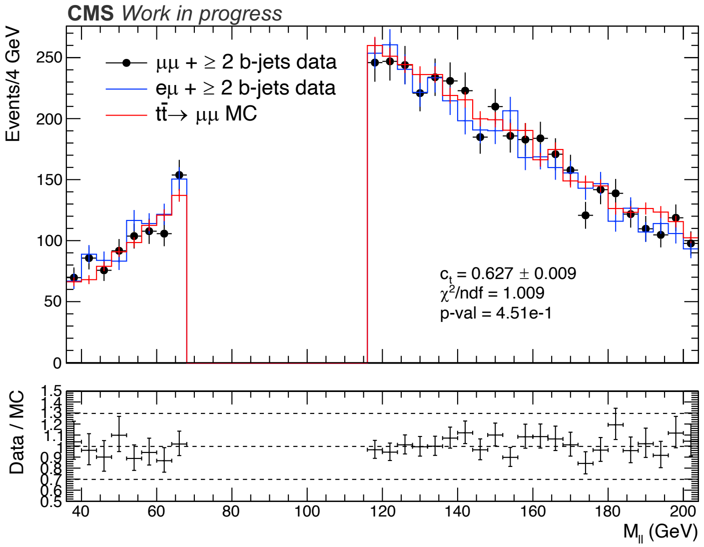
# Background estimation for Z + ≥ 2 b jets

- Data-driven method is used to estimate $t\bar{t}$ background, which is based on the $e\mu$ sample.
  - $t\bar{t}$ decay to $e\mu$ events are the dominant contribution.

$$N_{\ell\ell} = c_t \cdot N_{e\mu}, \quad \ell\ell = ee, \mu\mu \qquad \text{with } c_t \sim \frac{1}{2}$$

- $e\mu$ selection:
  - Isolated lepton passing the same selection criteria in the signal region.
  - MET $> 80$ GeV to enrich the $t\bar{t}$ events.
- Estimation of $c_t$:
  - Fit $e\mu$ sample to the sidebands of dilepton invariant mass ($40\ \text{GeV} < M_{ll} < 66\ \text{GeV}$ or $M_{ll} > 116\ \text{GeV}$) to obtain the normalization factor $c_t$




The $c_t$ coefficient extracted by $e\mu$ invariant mass fit.

| **Data Sample** | Z ($\rightarrow ee$) + $\geq$ 2 b jets | Z ($\rightarrow \mu\mu$) + $\geq$ 2 b jets |
|---|---|---|
| 2016 | $0.526 \pm 0.008$ | $0.627 \pm 0.009$ |
| 2017 | $0.530 \pm 0.007$ | $0.620 \pm 0.007$ |
| 2018 | $0.517 \pm 0.006$ | $0.621 \pm 0.006$ |

- $c_t$ obtained in muon channel is larger than which obtained in electron channel.

# Uncertainties in integrated cross section for Z + ≥ 2 b jets

- The dominant uncertainties are from JES,JER, MET measurement and b-tagging.
- Electron selection uncertainties (trigger, reconstruction and ID) is high in the electron channel, but it drops in combined results.

| Uncertainty (%) | $ee$ | $\mu\mu$ | $\ell\ell$ |
|---|---|---|---|
| Statistical | 7.7 | 5.9 | 4.6 |
| JES, JER | 6.9 | 5.4 | 5.8 |
| b tagging/mistagging | 5.4 | 6.0 | 5.8 |
| Unclustered energy of $p_T^{miss}$ | 3.5 | 3.7 | 3.6 |
| Background estimation | 2.3 | 2.4 | 2.4 |
| Pileup reweighting | 2.9 | 2.1 | 2.4 |
| Electron selection | 4.3 | – | 1.4 |
| Luminosity | 1.6 | 1.6 | 1.6 |
| Muon selection | – | 1.0 | 0.7 |
| L1 prefiring | 0.3 | 0.2 | 0.3 |
| $\mu_R$ and $\mu_F$ scales | 2.5 | 2.3 | 2.5 |
| PDF | 0.3 | 0.3 | 0.3 |
| $\alpha_s$ | 0.1 | 0.1 | 0.1 |

# Integrated Cross Section for Z + ≥ 2 b jets

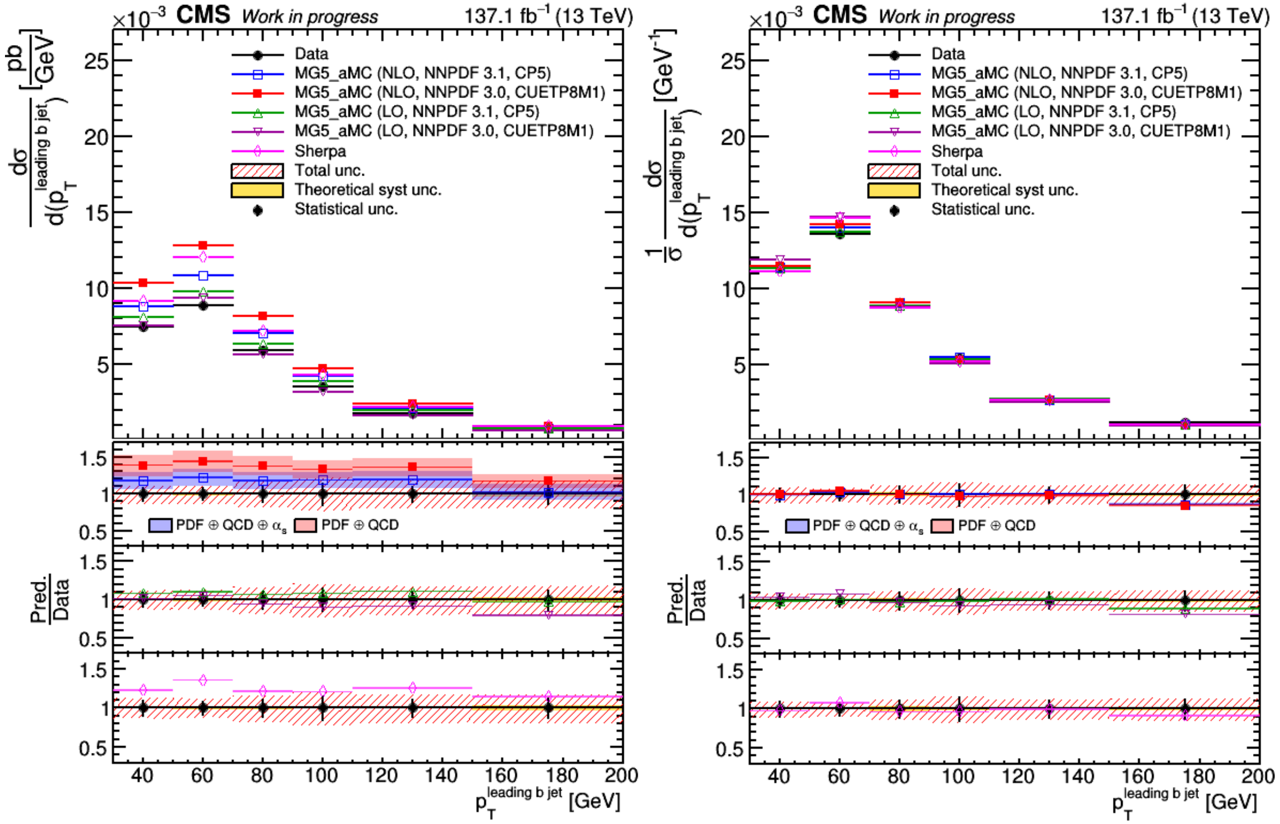
| Channel | Measured (pb) | MG5_aMC(LO) (pb) | | MG5_aMC(NLO) (pb) | | SHERPA (pb) |
|---|---|---|---|---|---|---|
| | | NNPDF 3.1 CP5 | NNPDF 3.0 CUETP8M1 | NNPDF 3.1 CP5 | NNPDF 3.0 CUETP8M1 | |
| $ee$ | $0.66 \pm 0.05 \pm 0.07 \pm 0.02$ | 0.72 | 0.62 | $0.77 \pm 0.07$ | $0.89 \pm 0.08$ | 0.84 |
| $\mu\mu$ | $0.65 \pm 0.04 \pm 0.06 \pm 0.02$ | 0.71 | 0.64 | $0.77 \pm 0.07$ | $0.91 \pm 0.09$ | 0.84 |
| $\ell\ell$ | $0.65 \pm 0.03 \pm 0.07 \pm 0.02$ | 0.71 | 0.63 | $0.77 \pm 0.07$ | $0.90 \pm 0.09$ | 0.84 |

- Z + ≥ 2 b jets cross sections is 0.65 ± 0.03 (stat) ± 0.07 (syst) ± 0.02 (theo) pb in combined channel.
- The measured integrated cross section is in good agreement with MADGRAPH5_aMC@NLO (MG5_aMC) at LO.
- The MG5_aMC (NLO, NNPDF 3.1, CP5), MG5_aMC (NLO, NNPDF 3.0, CUETP8M1) and SHERPA predictions overestimate the measured cross section values by 29%, 38%, and 21%.

MC predictions are normalized to NNLO calculation

# Unfolded differential & normalized differential cross section distributions

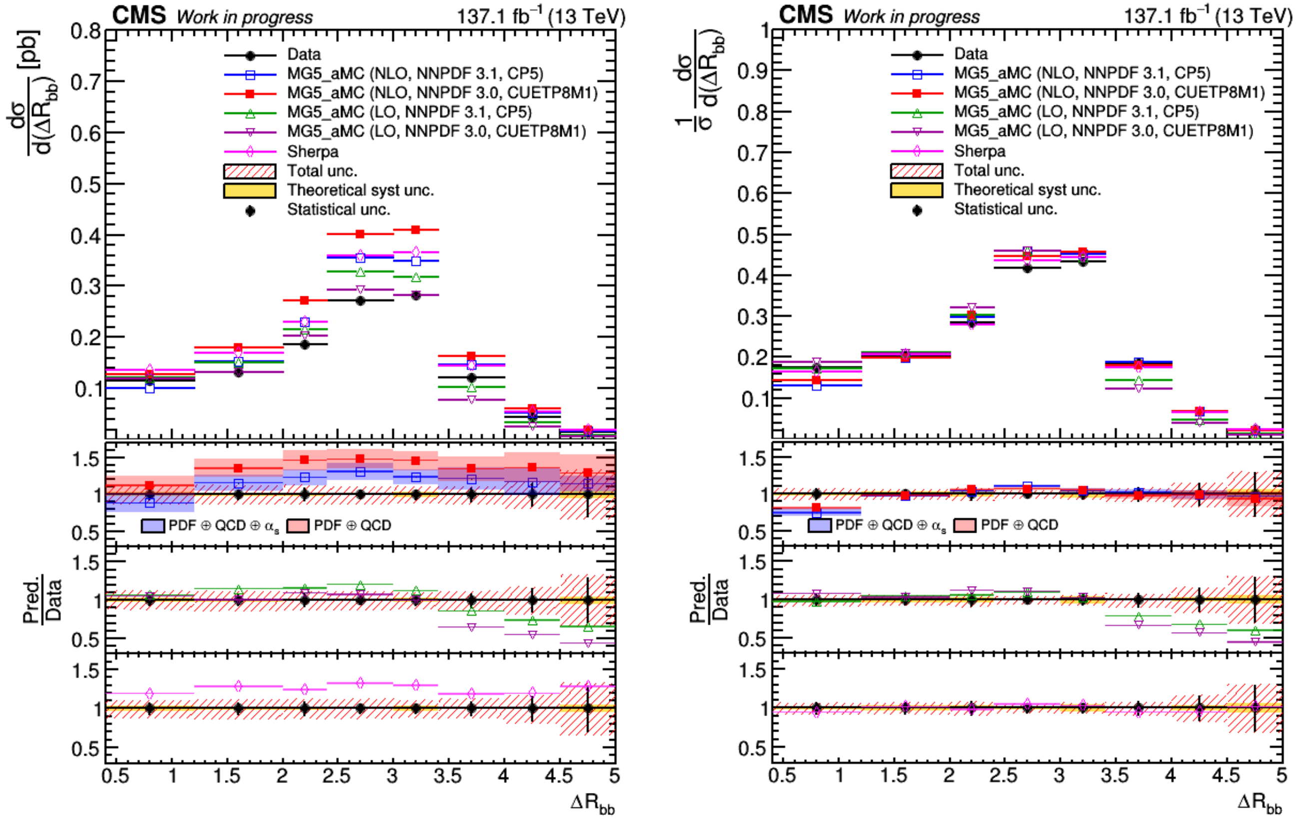
- Z + $\geq$ 2 b jets: leading jet $p_{\mathrm{T}}$




- The NLO and SHERPA samples overestimate the data ~20-50%.
- LO samples have a better agreement with data compared with NLO and SHERPA samples.
- Shape of data distribution is well described by all three MC simulations.

# Unfolded differential & normalized differential cross section distributions

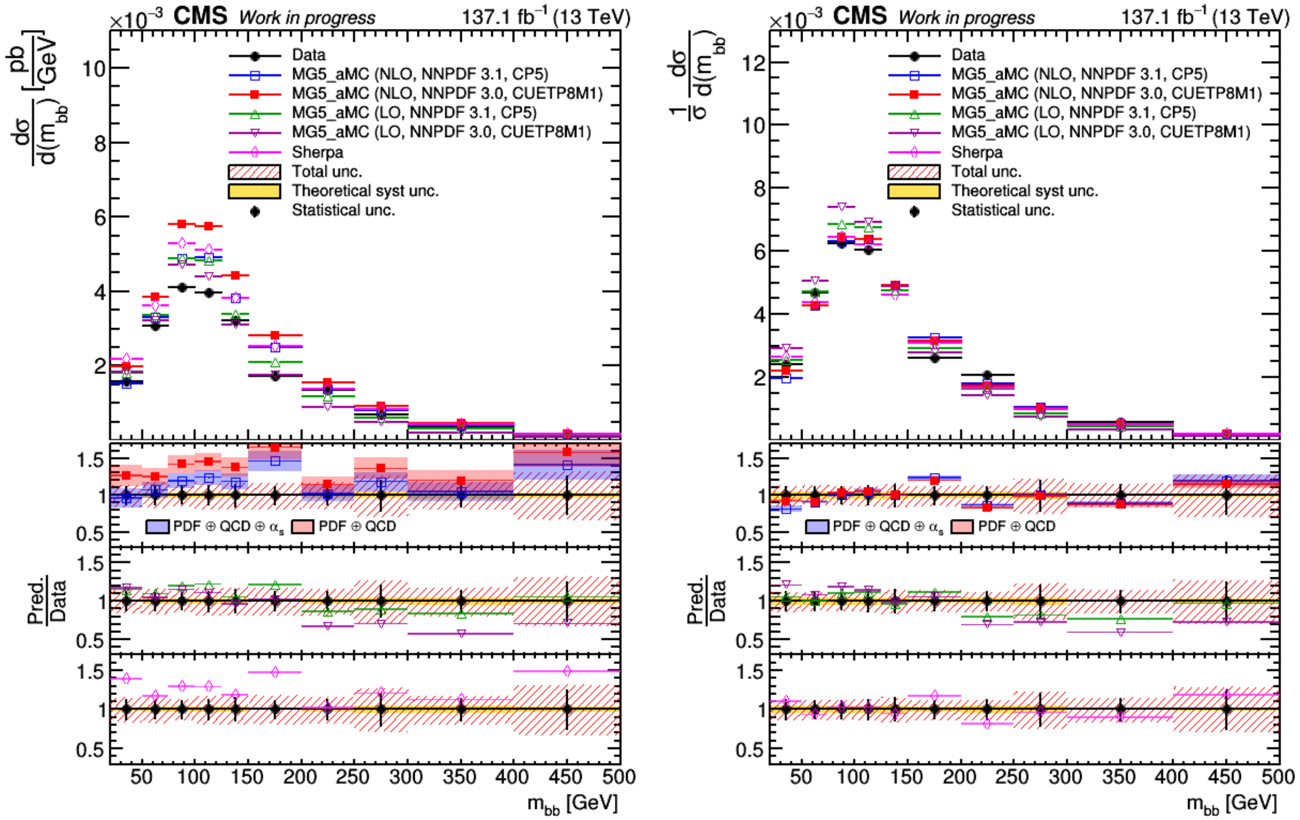
- Z + ≥ 2 b jets: $\Delta R_{bb}$




- NLO and SHERPA samples overestimate the data at high $\Delta R_{bb}$ region, while the LO samples have the opposite trend.
- Shape of data distribution is well described by SHERPA simulation.

# Unfolded differential & normalized differential cross section distributions

- Z + $\geq$ 2 b jets: $m_{bb}$




- NLO and SHERPA samples overestimates the data at high $m_{bb}$ region, while the LO (NNPDF 3.0, CUETP8M1) sample has the opposite trend.

# Summary

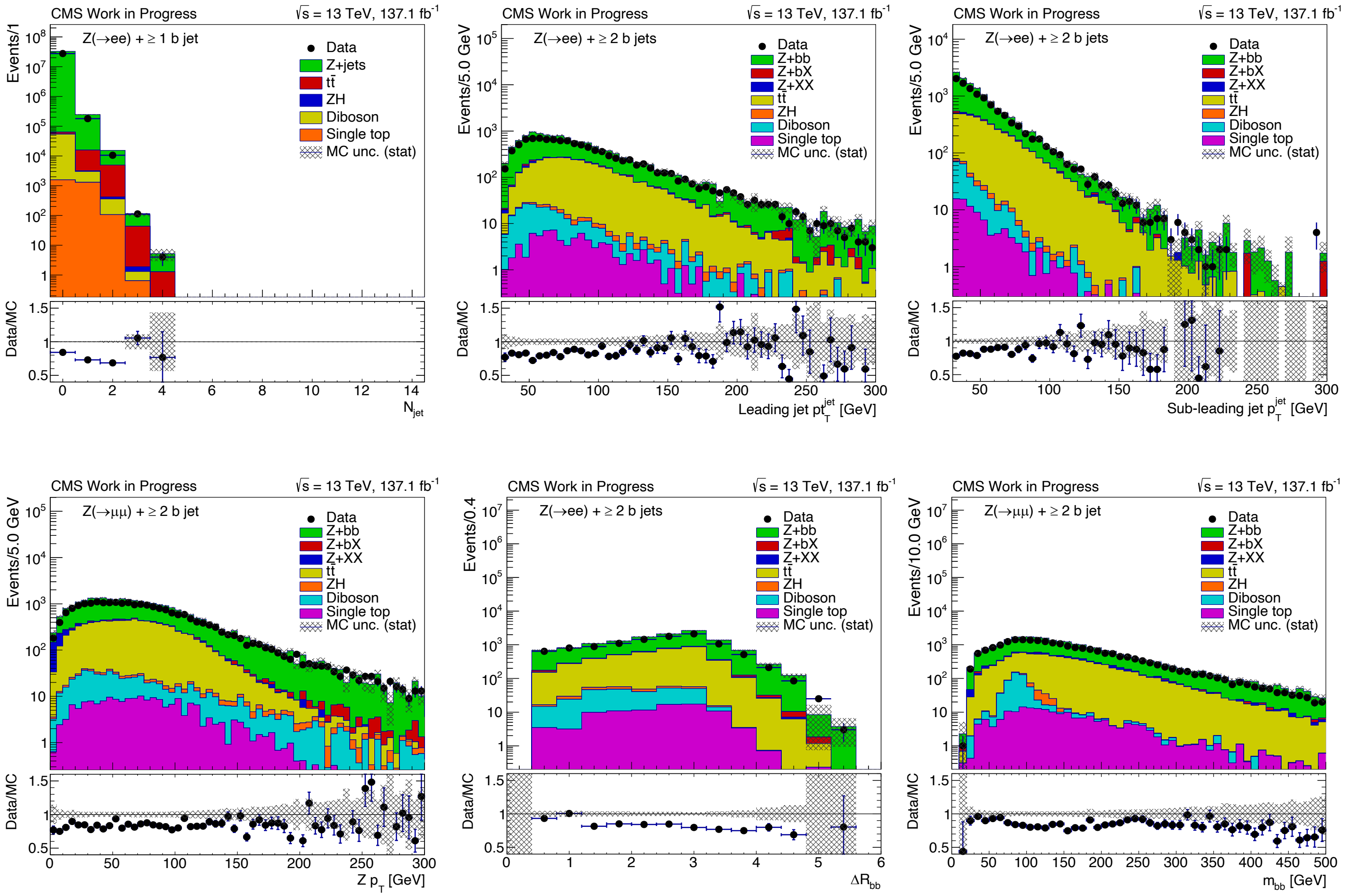
- Z + $\geq$ 2 b jets cross section:
  - 0.66 ± 0.05 (stat) ± 0.07 (syst) ± 0.02 (theo) pb in $ee$ channel
  - 0.65 ± 0.04 (stat) ± 0.06 (syst) ± 0.02 (theo) pb in $\mu\mu$ channel
  - 0.65 ± 0.03 (stat) ± 0.07 (syst) ± 0.02 (theo) pb in combined channel.
- Predictions of three different MC simulations (MG5_aMC NLO, MG5_aMC LO, SHERPA ) have been used to compare with data for a variety of distributions.
- MG5_aMC LO samples describe Z + $\geq$ 2 b jets events better, while MG5_aMC NLO and SHERPA samples usually overestimate the data rate.

# Backup

University at Buffalo The State University of New York

# Some distributions of Z + ≥ 2 b jet observables

- According to b jet multiplicity distribution, we have up to 4 b jets in some events.
- Data and MC difference in jet $p_T$ and dilepton $p_T$ distributions is ~10-20%.
- Data and MC difference in $\Delta R_{bb}$ distributions is ~20-30%.



# Unfolding

- The reconstructed distributions are unfolded to particle-level quantities using the TUnfold package (v17.5) to correct for the detector resolution and the selection efficiencies.
- The particle-level jets are formed from stable particles (c$\tau$ > 1 cm), except neutrinos and overlapping leptons from Z boson decays, using the same anti-$k_T$ algorithm used to reconstruct jets.
- If an event consists of at least two particle-level jets containing a b hadron, the event is classified as a Z + $\geq$ 2 b jets event.
- The generator-level leptons are "dressed" by adding the momentum of all photons within $\Delta R \leq 0.1$ around the lepton direction.
- The reconstructed distributions are unfolded with a response matrix that describes the migration probability between the particle- and reconstruction-level quantities.
- The response matrix is constructed when reconstruction-level leptons and b jet objects are spatially matched to the corresponding particle-level objects within $\Delta R < 0.3$.

# Uncertainties in differential cross section for Z + ≥ 2 b jets

| Observable/Uncertainty | $p_T^{b1}$ | $p_T^{b2}$ | $\eta^{b1}$ | $p_T^Z$ | $\Delta R_{bb}$ | $\Delta R_{Zbb}^{min}$ | $A_{Zbb}$ | $M_{Zbb}$ | $M_{bb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Statistical | 8.6–15.5 | 7.0–37.6 | 6.2–11.9 | 6.2–25.3 | 5.6–29.2 | 5.3–43.5 | 4.5–25.0 | 8.1–22.1 | 11.3–25.9 |
| JES, JER | 6.0–12.8 | 5.1–14.3 | 6.8–8.2 | 5.6–14.8 | 5.9–14.9 | 4.5–23.8 | 4.3–13.2 | 4.8–13.7 | 4.2–17.5 |
| b tagging/mistagging | 4.4–7.7 | 4.3–9.2 | 4.3–6.2 | 4.1–13.7 | 3.5–6.4 | 3.8–14.5 | 5.3–7.3 | 4.2–7.2 | 4.5–6.6 |
| Unclustered energy of $p_T^{miss}$ | 1.3–6.8 | 2.1–8.0 | 2.7–4.2 | 3.1–6.8 | 1.1–5.0 | 0.7–4.8 | 2.1–4.8 | 2.1–5.1 | 1.3–8.4 |
| Background simulation | 0.6–3.7 | 0.8–4.9 | 1.0–2.9 | 0.6–3.8 | 0.5–3.0 | 0.5–3.8 | 1.9–3.1 | 0.6–4.1 | 1.1–3.5 |
| Pileup reweighting | 1.4–4.1 | 1.6–6.8 | 1.7–2.7 | 1.6–3.3 | 1.9–3.1 | 0.9–3.9 | 1.7–4.3 | 1.7–3.2 | 1.5–3.5 |
| Electron selection | 0.9–1.7 | 1.1–2.0 | 1.2–1.5 | 0.9–1.7 | 0.7–1.5 | 0.5–1.8 | 1.1–1.6 | 1.0–1.8 | 0.9–1.6 |
| Muon selection | 0.5–0.8 | 0.5–0.9 | 0.7–3.3 | 0.4–0.9 | 0.5–0.8 | 0.5–0.8 | 0.6–0.8 | 0.5–0.9 | 0.6–0.8 |
| $\mu_R$ and $\mu_F$ scales | 0.5–3.2 | 0.7–4.6 | 1.8–4.7 | 0.7–4.5 | 0.6–4.1 | 0.7–4.7 | 1.3–4.9 | 1.7–3.8 | 1.1–4.8 |
| PDF | 0.5–0.8 | 0.5–1.4 | 0.5–0.7 | 0.5–1.1 | 0.5–2.5 | 0.5–4.0 | 0.5–0.9 | 0.5–0.8 | 0.5–0.9 |
| $\alpha_s$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.5–1.6 | 0.5–0.9 | 0.5 | 0.5 | 0.5 |

- $b_1$ ($b_2$) is the leading (sub-leading) b-jet.
- statistical uncertainty is the dominant uncertainty.
- JES, JER, MET measurement and b-tagging are major contributions of uncertainties

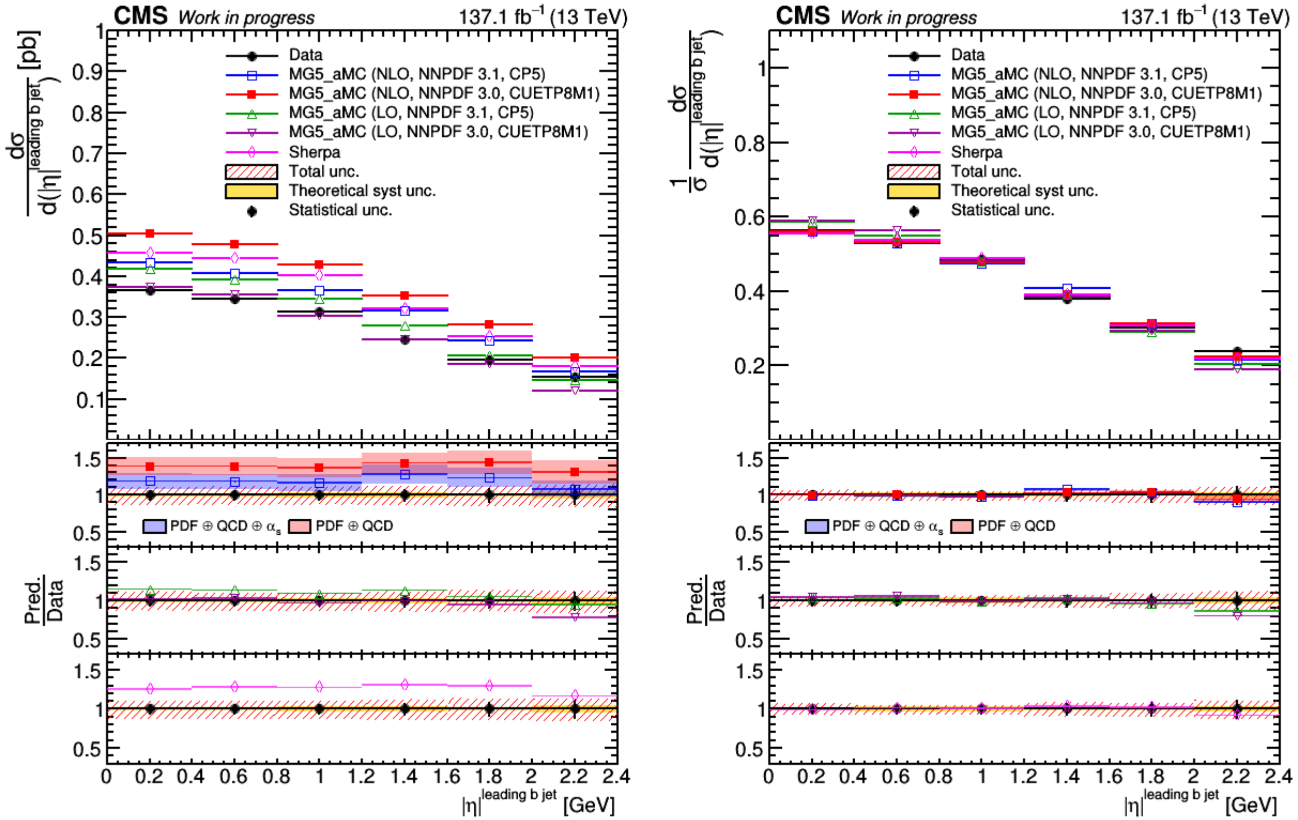
# Uncertainties in normalized differential cross section for Z + ≥ 2 b jets

| Observable/Uncertainty (%) | $p_T^{b1}$ | $p_T^{b2}$ | $\eta^{b1}$ | $p_T$ | $\Delta R_{bb}$ | $\Delta R_{Zbb}^{min}$ | $A_{Zbb}$ | $M_{Zbb}$ | $M_{bb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Statistical | 9.7–16.2 | 8.3–37.9 | 7.7–12.8 | 7.7–25.7 | 7.2–29.6 | 7.0–43.7 | 6.4–25.4 | 9.3–22.6 | 12.2–26.3 |
| JES, JER | 3.2–7.7 | 1.6–9.4 | 0.8–1.3 | 0.9–7.9 | 0.8–7.9 | 1.0–17.5 | 0.9–6.1 | 2.2–8.1 | 3.5–12.5 |
| b tagging/mistagging | 0.5–2.2 | 0.7–3.8 | 0.6–1.7 | 0.5–8.4 | 0.5–2.4 | 0.5–9.5 | 0.5–1.6 | 0.5–1.7 | 0.5–1.3 |
| Unclustered energy of $p_T^{miss}$ | 0.5–3.1 | 0.8–7.1 | 0.5–0.9 | 0.5–3.0 | 0.5–2.5 | 0.5–6.1 | 0.5–3.9 | 0.7–3.1 | 0.5–4.6 |
| Background simulation | 0.5–1.8 | 0.5–2.7 | 0.5–1.4 | 0.5–1.9 | 0.5–2.1 | 0.5–2.3 | 0.5–0.7 | 0.5–1.8 | 0.5–1.3 |
| Pileup reweighting | 0.5–1.8 | 0.5–4.3 | 0.5–0.8 | 0.5–2.7 | 0.5–0.8 | 0.5–1.7 | 0.5–1.7 | 0.5–1.1 | 0.5–1.3 |
| Electron selection | 0.5 | 0.5–0.7 | 0.5 | 0.5 | 0.5–0.6 | 0.5–1.0 | 0.5 | 0.5 | 0.5 |
| Muon selection | 0.5 | 0.5 | 0.9–2.3 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| $\mu_R$ and $\mu_F$ scales | 0.5–2.9 | 0.6–4.6 | 1.4–3.9 | 1.4–3.8 | 1.3–5.0 | 1.1–4.8 | 1.7–4.6 | 2.0–4.7 | 1.3–4.9 |
| PDF | 0.5–0.6 | 0.7–1.3 | 0.5–0.6 | 0.5–1.0 | 0.5–2.1 | 2.6–3.2 | 0.5–0.9 | 0.5–0.8 | 0.5–0.8 |
| $\alpha_s$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.5–1.3 | 0.5–0.6 | 0.5 | 0.5 | 0.5 |

- $b_1$ ($b_2$) is the leading (sub-leading) b-jet.
- statistical uncertainty is the dominant uncertainty.

# Unfolded differential & normalized differential cross section distributions

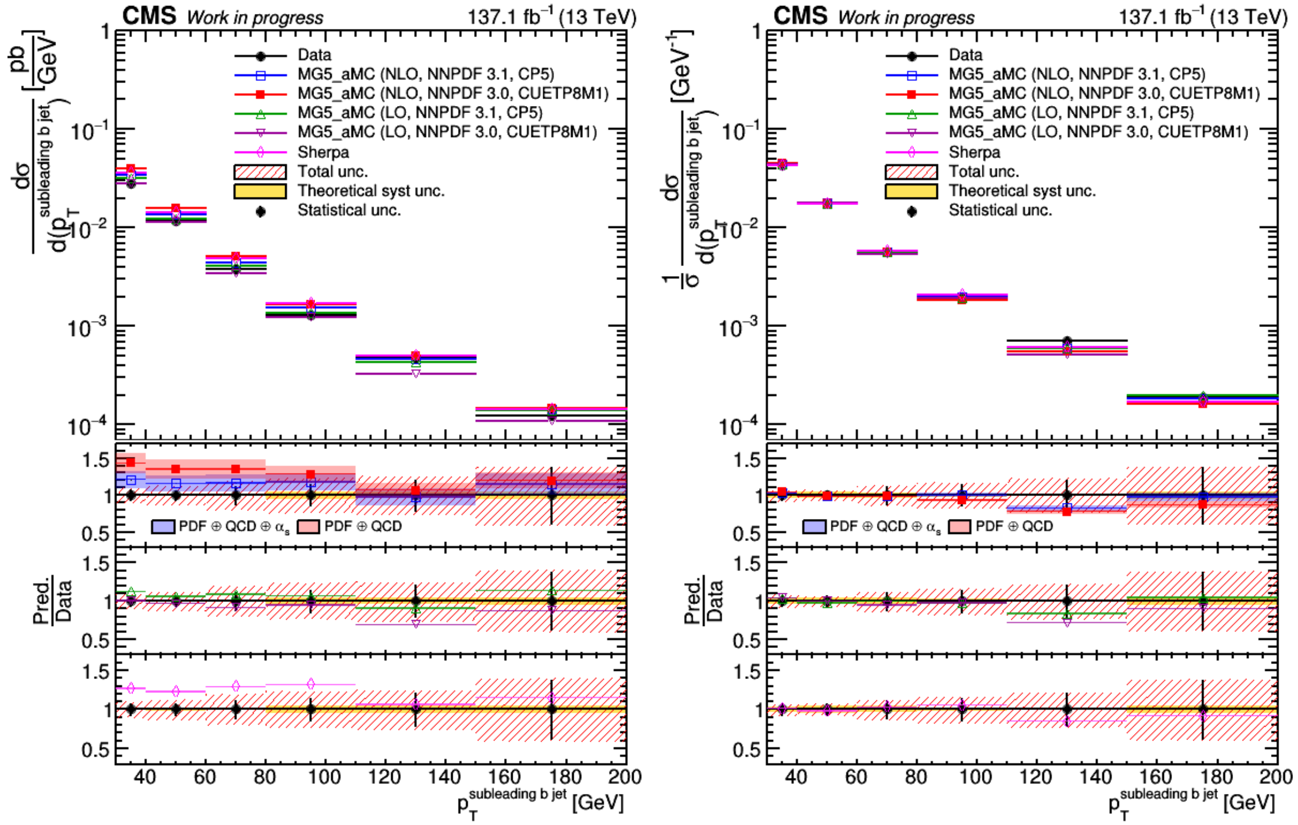
- Z + ≥ 2 b jets: leading jet absolute pseudorapidity




- Shape of data distribution is well described by all three MC simulations.
- LO samples have a better agreement with data compared with NLO and SHERPA samples.

# Unfolded differential & normalized differential cross section distributions

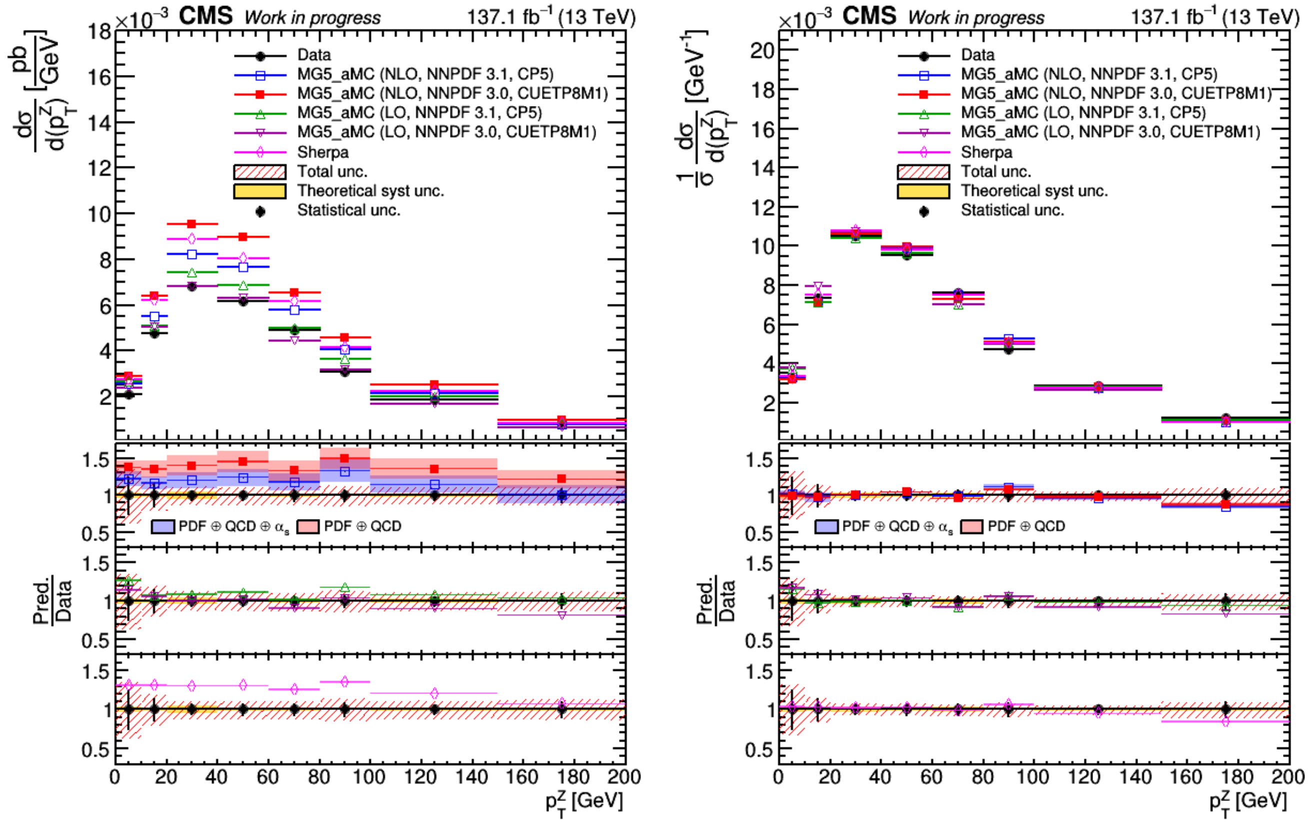
- Z + ≥ 2 b jets: sub-leading jet $p_T$




- Shape of data distribution is well described by all three MC simulations besides $110 < p_T < 150$ GeV region.

# Unfolded differential & normalized differential cross section distributions

- Z + $\geq$ 2 b jets: dilepton $p_{\mathrm{T}}$




- Besides high $p_{\mathrm{T}}$ region, shape of data distribution is well described by all three MC simulations.