Z boson production in association with at least two b jets in pp collisions at 13 TeV: signal versus background

Peter Young
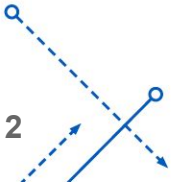
State University of New York at Buffalo

New Perspectives - August 18, 2021

University at Buffalo The State University of New York

Overview

- Motivation

- Signal vs. Background

- Data-Driven Background Estimation

- Analysis Results

- Summary

Motivation (I)

- Z boson produced in associated with bottom quark jets (b-jets) provide important tests of perturbative quantum chromodynamics (pQCD) calculations

- The processes form major background for a variety of physics analyses
    - Higgs production in association with a Z boson, ZH(H➡bb/cc)
    - Beyond Standard Model (BSM) search for new generations of heavy quarks (b',t') decaying to Z boson and a b quark (Z+b jet)

- Measurements of various event kinematic properties are important to tune Monte Carlo (MC) simulations of Z + jets production
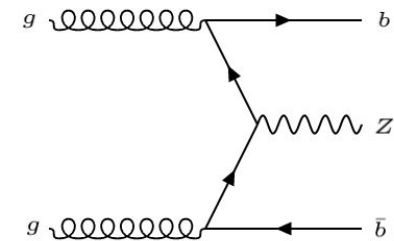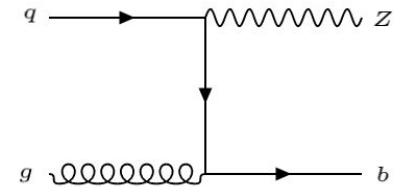

Fig 1. Feynman diagrams of tree-level contributions for Z + ≥ 1 b-jet (top) and Z + ≥ 2 b-jets (bottom) processes

3

Motivation (II)

- Z boson produced in association with ≥ 2 b-jets

  - Z boson decays into lepton (e^+e^-, $\mu^+\mu^-$) pair

  - Expect at least 2-jets, tagged by DeepCSV as b-jets

- Part of first on-going CMS analysis of Z+b-jets done at $\sqrt{s}$ = 13 TeV using full Run II data, 137 fb^{-1}

- Previous Z + heavy-flavor (HF) jets done at $\sqrt{s}$ = 8 TeV, Run I data*
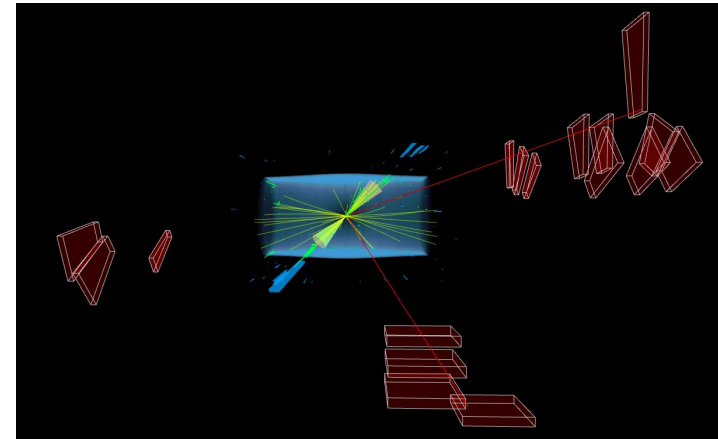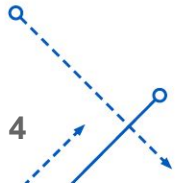

Fig 2. Event candidate for Z + two b-jets, Z decay into $\mu^+\mu^-$ pair

*Eur. Phys. J. C **77** (2017) 751, arXiv:1611.06507v2 [hep-ex].

Event Selection Criteria (I)

- Full Run II data with integrated luminosity of 137.1 fb^{-1} (2016, 2017, 2018)

- Trigger: SingleElectron, SingleMuon Trigger

- Leptons: isolated with leading (sub-leading) lepton $p_T > 35$ (25) GeV $| \eta | < 2.4$

- Z boson: pair of oppositely charged leading and sub-leading leptons within the mass window
        71 GeV < $M_{\ell\ell}$ < 111 GeV

- Jet: jets reconstructed with anti-k_T algorithm with distance parameter of 0.4
    $p_T > 30$ GeV, $| \eta | < 2.4$, lepton overlap removal, tight jet ID
    jets from pileup events are excluded
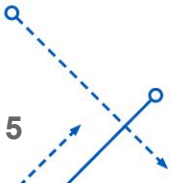
Event Selection Criteria (II)

- b-jets: jets tagged with DeepCSV tagger

  tight tagging operating point is required to reduce Drell-Yan (DY) + jets background:
  DY+bX, DY+XX, where X = c- or light jets
  loose, medium, tight operating points correspond to approximately 80%, 60%, 50%
  b quark tagging efficiency and 10%, 1%, and 0.1% misidentification rate for light jets

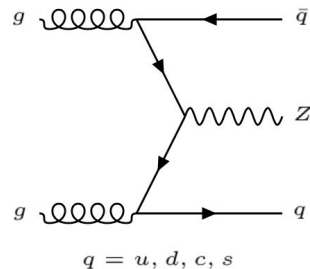- Missing Transverse Energy (MET), MET < 50 GeV → reduces tt background

Fig 3. Background process for Z + ≥ 2 b-jets where c- or light (u,d,s) jets are produced in association

6

Events after Z + ≥ 2 b-jets Selection

- Numerous backgrounds in signal (DY+bb) region

- tt background is not negligible, accounts for ~⅓ of total events

- tt accounts for ~85% of overall background events

Table 12: Number of events after Z + ≥ 2 b jets selection

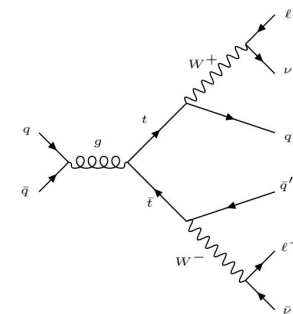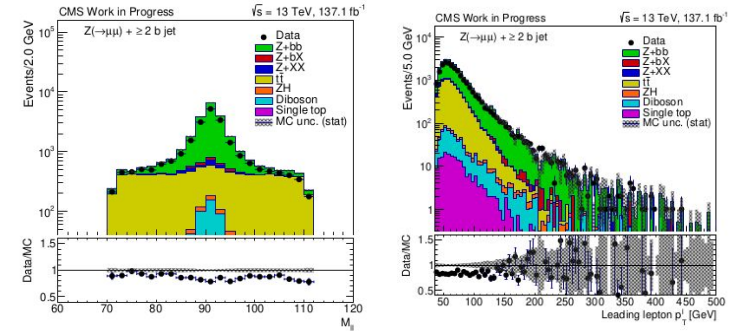|	2016		2017		2018	
	Electron	Muon	Electron	Muon	Electron	Muon
DY+bb	1976	4206	2241	4557	4361	8934
DY+bX	38	89	29	81	72	190
DY+XX	10	89	2	55	43	237
tt	784	1748	1102	2362	1825	3865
single t	18	40	22	59	48	99
VV	44	97	65	136	98	218
ZH	11	22	13	27	23	45
MC	2880	6291	3473	7278	6470	13589
Data	2549	5503	3070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

Fig 4. (Left to right, top to bottom) Dilepton invariant mass and leading lepton p_T distributions for muon channel (combined Run II data), Feynman diagram for tt production process

Signal vs. Background

- MET in tt and signal events are very different

$$t\bar{t} \rightarrow W^+ W^- b\bar{b} \rightarrow \ell^+ \ell^- \nu \bar{\nu} b\bar{b}$$

$$Z b\bar{b} \rightarrow \ell^+ \ell^- b\bar{b}$$
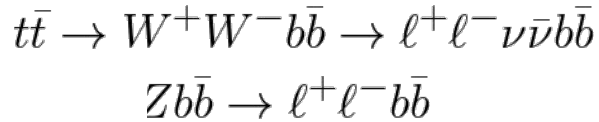
- Determine optimal MET cut to optimize event selection

- 50 GeV cut → 90% signal and 30% background events
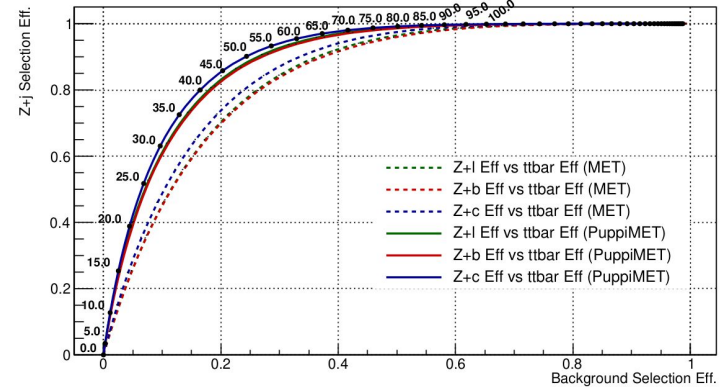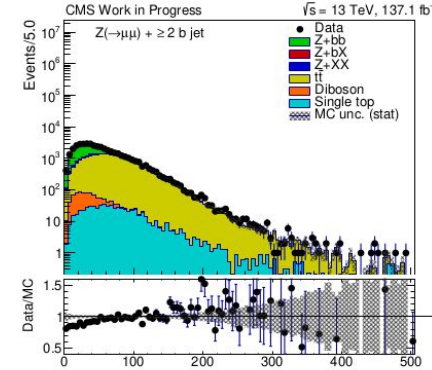 are kept

- Keep events where MET < 50 GeV

Fig 5. MET distribution (top) and Receiving operating characteristic (ROC) curve (bottom) for Z+jet and tt selection efficiency as a function of MET

8

Data-Driven Background Estimation (I)

- Backgrounds can be estimated using MC simulations

- tt can be more precisely estimated with data-driven methods

- Technique applied in earlier analyses at CDF (Tevatron)*, CMS**

- Outside Z signal region, events are mostly composed of background events
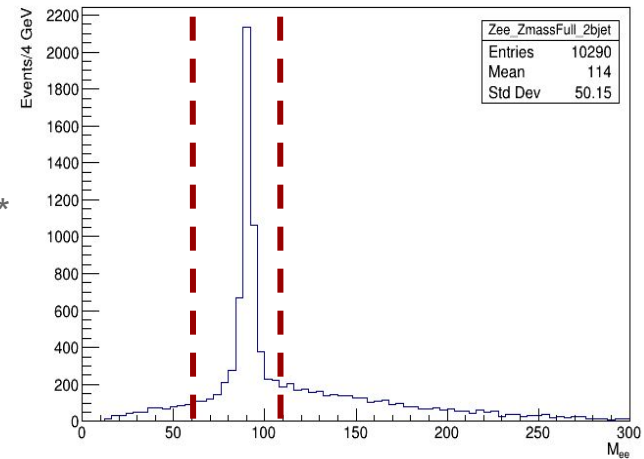
- Estimate background rate using sideband region

Fig 6. e^+e^- invariant mass distribution for 2018 data; the region within the dashed lines indicates the signal region

* *Phys. Rev. D* **79** (Mar, 2009) 052008, doi:10.1103/PhysRevD.79.052008.

***Eur. Phys. J. C* **77** (2017) 751, arXiv:1611.06507v2 [hep-ex].

Data-Driven Background Estimation (II)

- Use same trigger, event selection criteria as analysis

- Sideband regions $40 < M_{\ell\ell} < 66$ GeV $M_{\ell\ell} > 116$ GeV

- Enrich tt sample (suppress Z+jet events) by using events where MET > 80 GeV

- Match sideband regions (ee,μμ) with eμ events

- Fit dilepton data distribution to pull normalization constant between dilepton and eμ distributions

- Expect $c_t = 0.5$ due to combinatorics of final state particles

$$N_{\ell\ell} = c_t \cdot N_{e\mu} \qquad \ell\ell = ee, \mu\mu$$

$$c_t \sim \tfrac{1}{2}$$
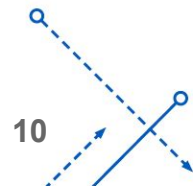
Results (I)

- Scale eμ sample by extracted coefficient
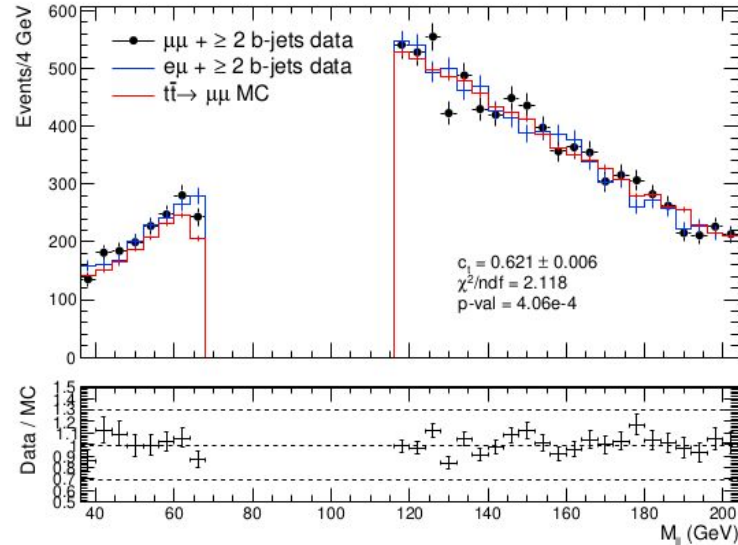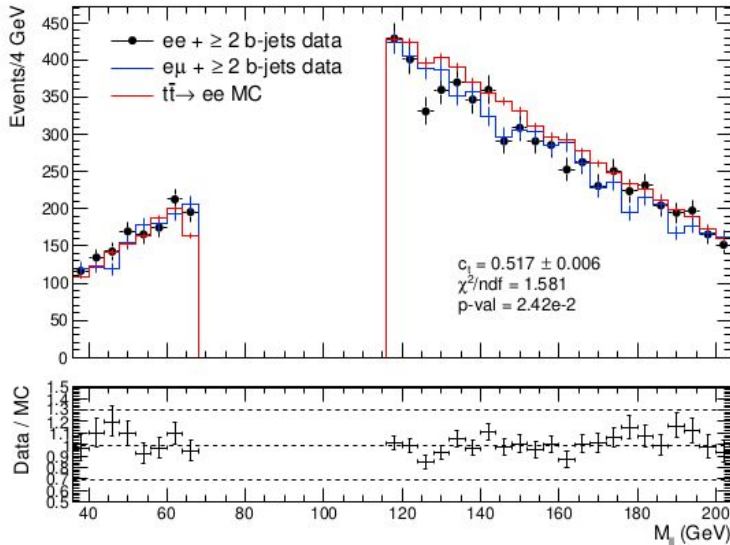- Scaled sample matches sideband regions and tt MC simulation within uncertainty

FIG 7. Z + ≥ 2 b-jets dilepton invariant mass sideband region fitted for electron (left) and muon (right) channels, 2018 data

Results (II)

- Subtract eμ events from the dilepton invariant mass distribution to isolate the signal
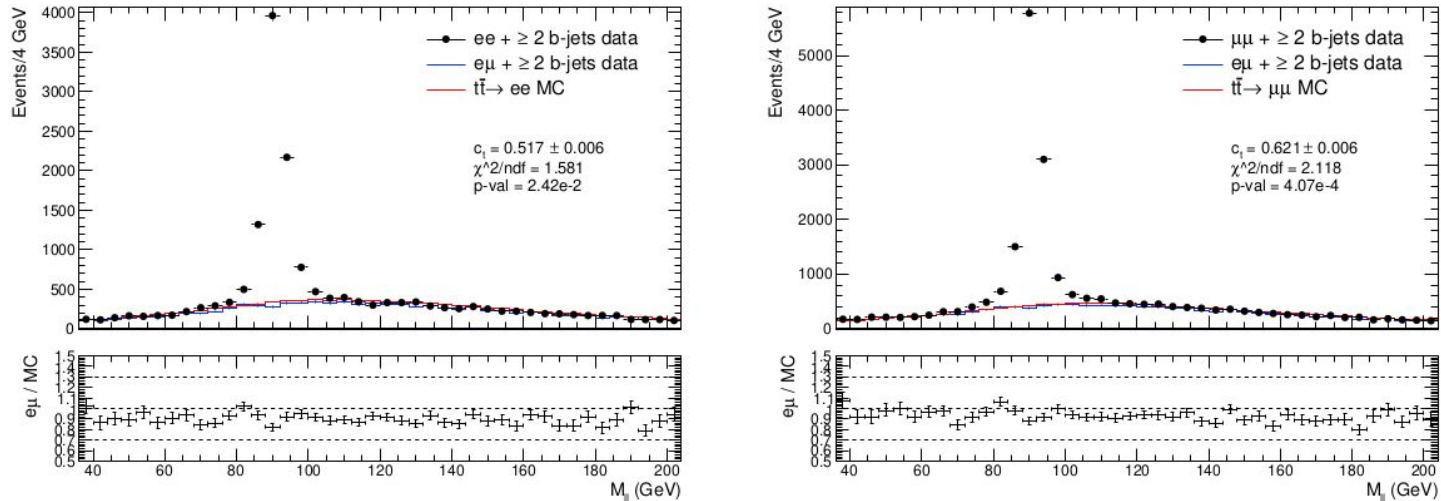- Scaled eμ sample matches tt MC over entire region within 10%

FIG 8. Z + ≥ 2 b-jets dilepton invariant mass fit between eμ data and ttbar MC samples for electron (left) and muon (right) channels, 2018 data

Results (III)

- Same process can be applied to MET, MET Significance distributions
- Various fits match within a maximum variation of 7%

$$S = \frac{E_T^{\mathrm{miss}}}{\sigma}$$

Sample	Z Mass	MET	MET Significance
2016 Electron Channel	0.526 ± 0.008	0.528 ± 0.006	0.524 ± 0.007
2017 Electron Channel	0.530 ± 0.007	0.559 ± 0.005	0.544 ± 0.007
2018 Electron Channel	0.517 ± 0.006	0.537 ± 0.004	0.519 ± 0.005

| Sample	Z Mass	MET	MET Significance
2016 Muon Channel	0.627 ± 0.009	0.643 ± 0.006	0.632 ± 0.007
2017 Muon Channel	0.620 ± 0.007	0.669 ± 0.005	0.640 ± 0.007
2018 Muon Channel	0.621 ± 0.006	0.656 ± 0.004	0.641 ± 0.005

13

Cross Section Results

- After Z + ≥ 2 b-jet selection and data-driven background estimation,

CHANNEL	MEASURED (pb)	MG5_aMC (LO) (pb)
Electron	0.66 ± 0.05 (stat) ± 0.08 (syst) ± 0.02 (theo)	0.677 ± 0.04 ± 0.01
Muon	0.65 ± 0.04 (stat) ± 0.07 (syst) ± 0.02 (theo)	0.677 ± 0.04 ± 0.01
Combined	0.65 ± 0.03 (stat) ± 0.07 (syst) ± 0.01 (theo)	0.677 ± 0.04 ± 0.01

- Measurements and predictions agree with each other
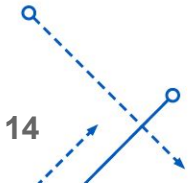
Summary

- Z+HF production provides important tests of pQCD calculations. It is an important background for numerous processes such as associated Higgs boson production ZH(H➜bb/cc) and searches for New Physics

- In Z + 2 b-jets production, the background is dominated by tt events which can be reduced by MET requirements.

- Background rate can be estimated more accurately with data-driven method.

- Overall results show agreement between MC and data.

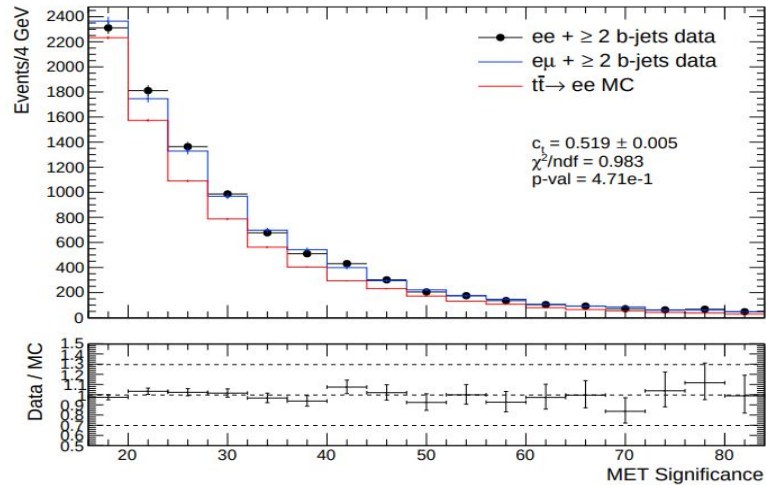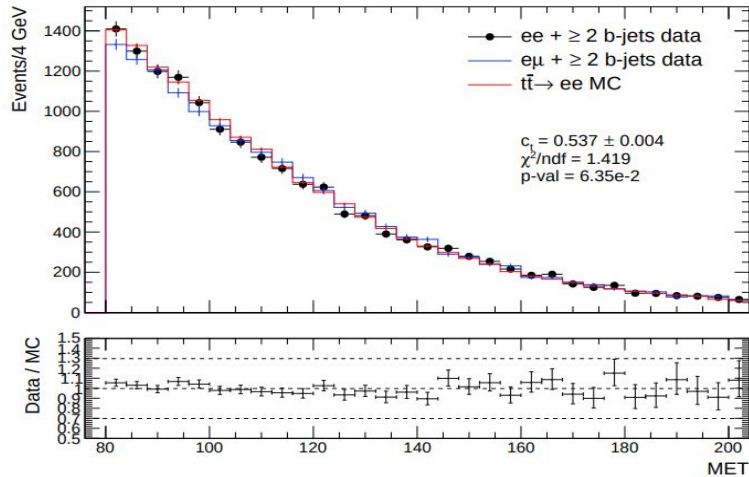Thank you for listening.

ADDITIONAL SLIDES

Additional Fits

FIG 9. Z + ≥ 2 b-jets MET distribution (left) and MET significance distribution (right) fit between eμ data and ttbar MC samples for electron channel, 2018 data

tt Production

- Top quark decays into W boson and quark

- W boson decays into charged lepton and neutrino

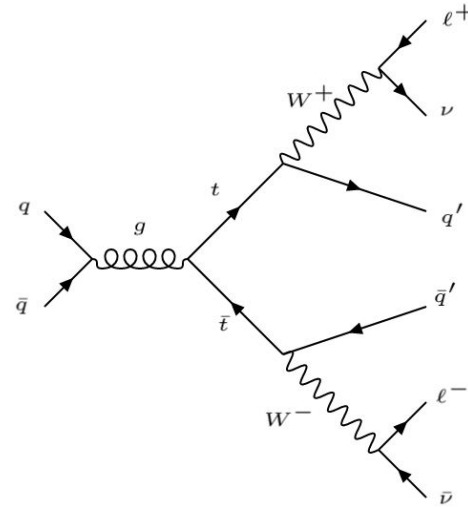- $\sigma(tt \rightarrow \ell\ell + MET + \geq 2$ jets$) \sim 170$ pb

FIG 10. Feynman diagram for ttbar production

Missing Transverse Energy

- In ideal case,

$$\sum_{i=0}^{N} E_{T,i} = 0$$

- Neutrinos are not detected by CMS detector, leaving an imbalance in transverse energy

- Random noise in system, errors in detectors can lead to some level of imbalance

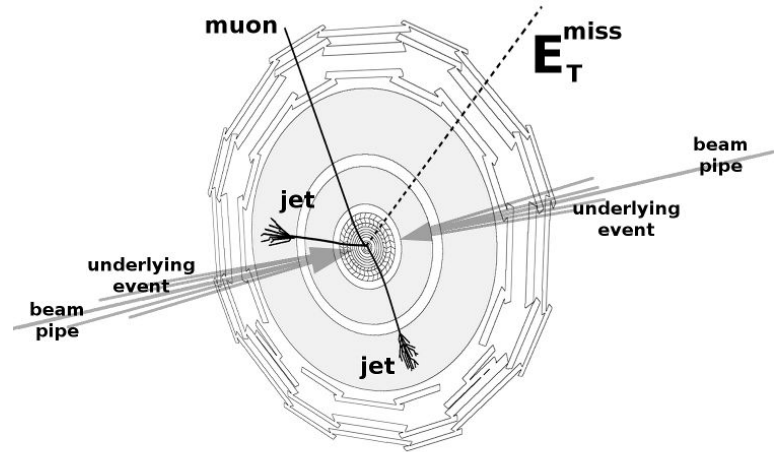- Very few events will truly have MET = 0 GeV

FIG 10. A sample cross section demonstration of missing transverse energy