Experimental overview of heavy flavor jets at LHCb

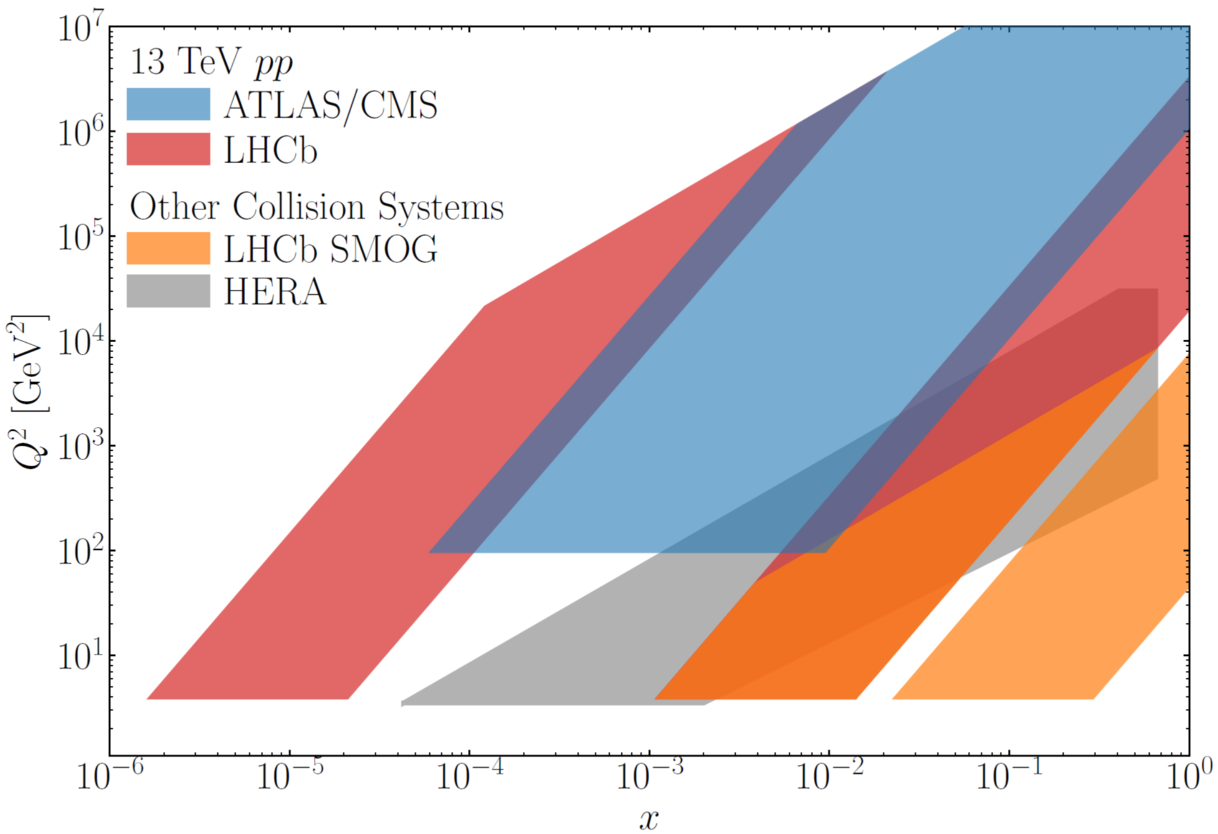
Kara Mattioli
University of Michigan

On behalf of the LHCb Collaboration

Jets at LHC Workshop

June 1, 2021

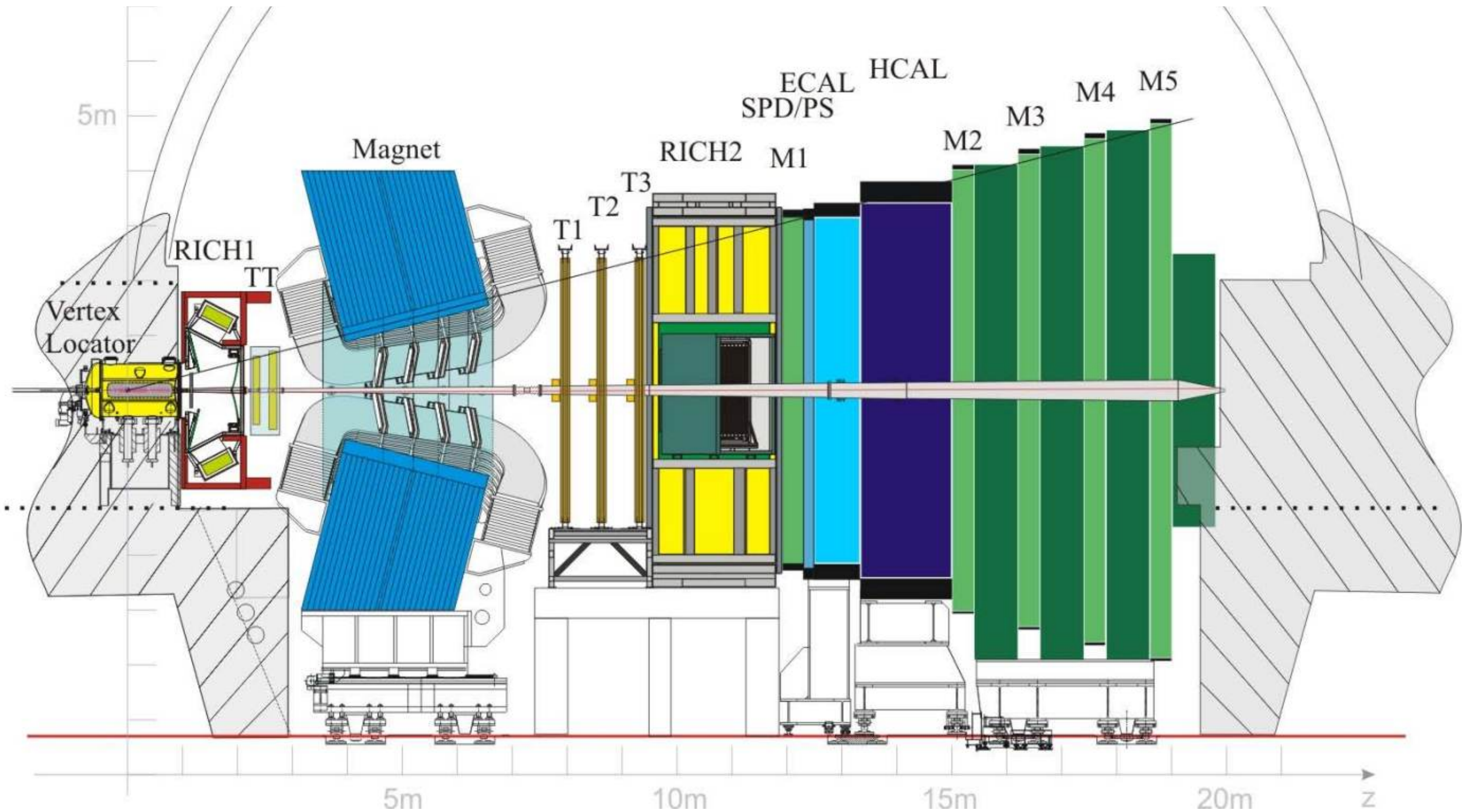
Heavy flavor jets in the forward region



- LHCb's forward acceptance (2< η < 5) allows it to access a complementary region of phase space to ALICE, ATLAS, and CMS
  - Extends region of phase space for tests of perturbative QCD calculations
- Measurements with heavy flavor jets in the forward region can constrain PDFs at high and low x
  - Typically where PDF uncertainties are the largest

The Large Hadron Collider beauty (LHCb) Detector

Forward spectrometer designed to study the production and decays of heavy flavor hadrons



Momentum resolution:
0.5% at low momentum to 1% at 200 GeV/c

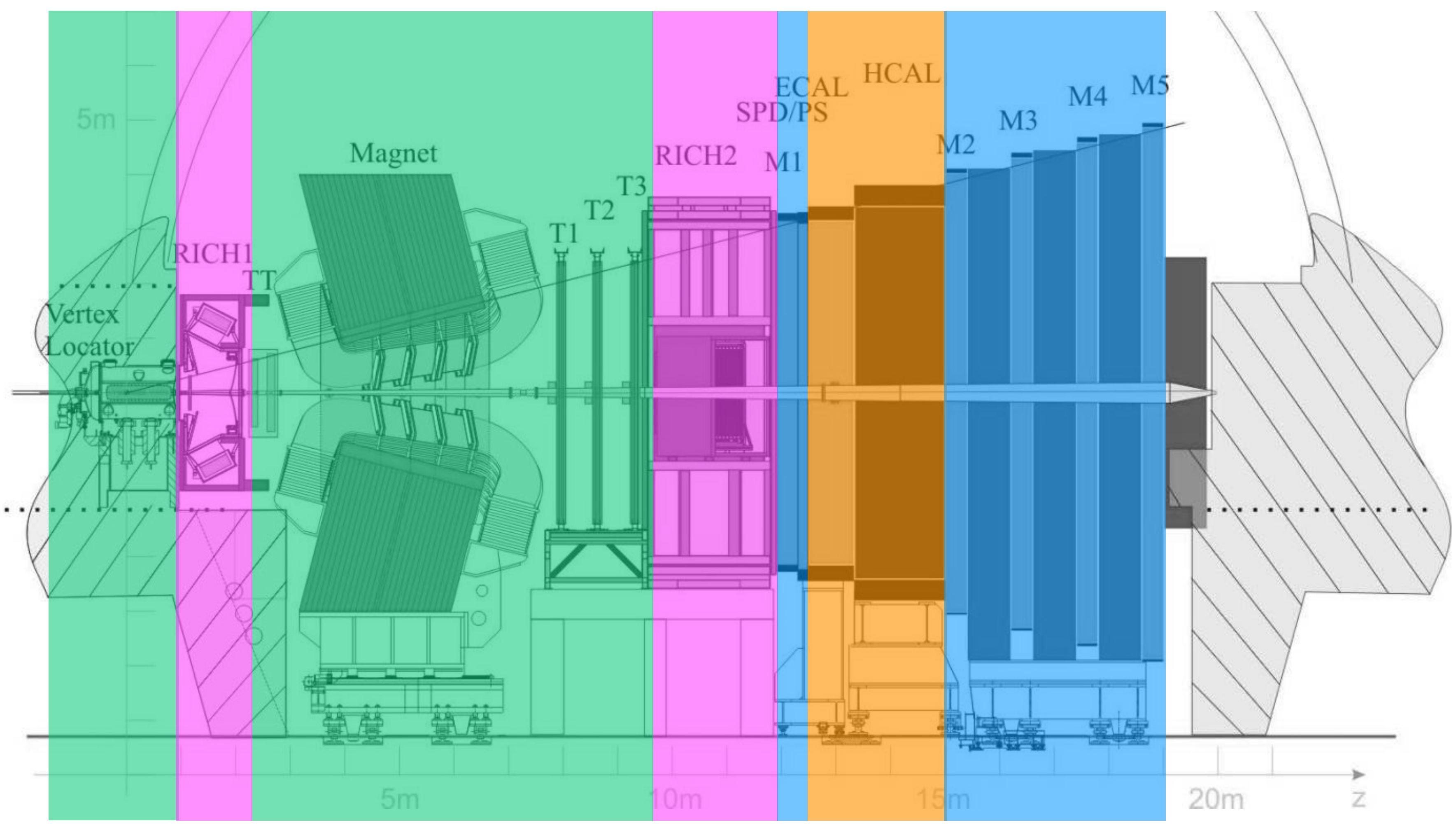
Impact parameter resolution:
15+29/p_T μm

JINST 3, S08005 (2010)

Int. J. Mod. Phys. A 30, 1530022 (2015)

The Large Hadron Collider beauty (LHCb) Detector

Full **hadronic and electromagnetic calorimetry**, **tracking**, **particle identification**, and **muon ID** in $2 < \eta < 5$



Jet reconstruction:

- Particle flow algorithm to select charged + neutral inputs
- Anti-k_T clustering with R=0.5

JINST 3, S08005 (2010)

Int. J. Mod. Phys. A 30, 1530022 (2015)

This talk:

1. Heavy flavor jet tagging at LHCb

2. Heavy flavor jet measurements at LHCb

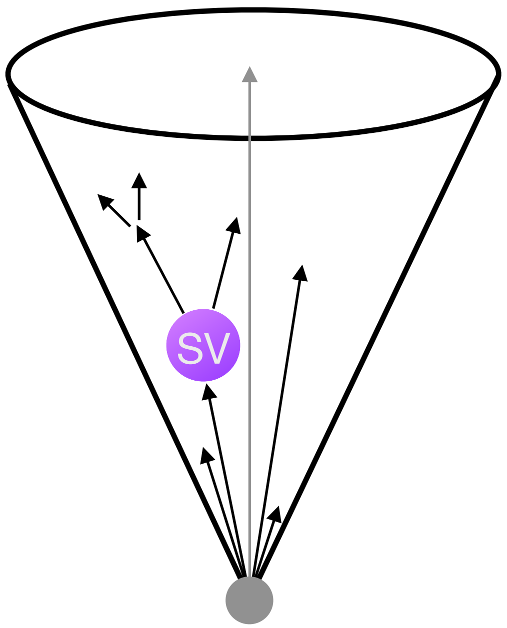
- Heavy flavor dijet measurements
- Top production measurements
- (W/Z) + heavy flavor jet measurements

Heavy flavor jet tagging at LHCb

Heavy flavor jet tagging at LHCb

Consists of two components:

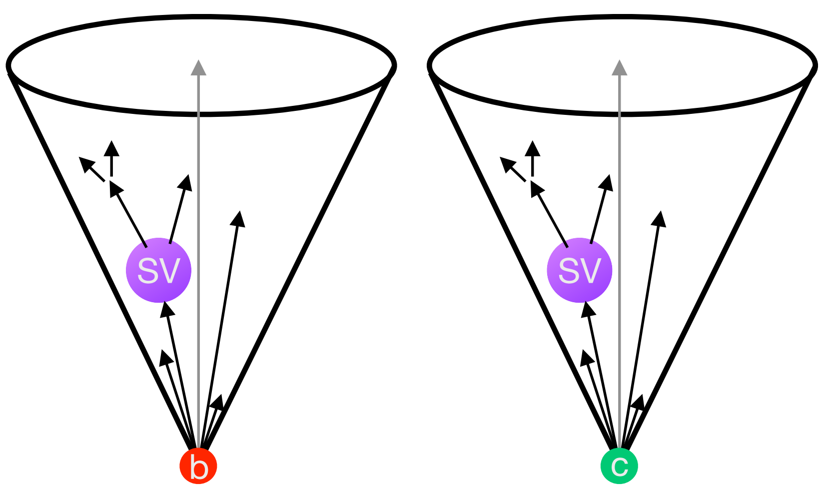
1. Secondary vertex tagging ("SV-tagging") algorithm



- Uses tracks inside and outside the jet with a significant p_T and PV displacement to construct all possible 2-track SVs, then merges SVs that share tracks into N-track SVs
- A jet is tagged as a heavy flavor jet if it has a SV passing all quality cuts within the jet cone (ΔR (SV, jet) < 0.5)

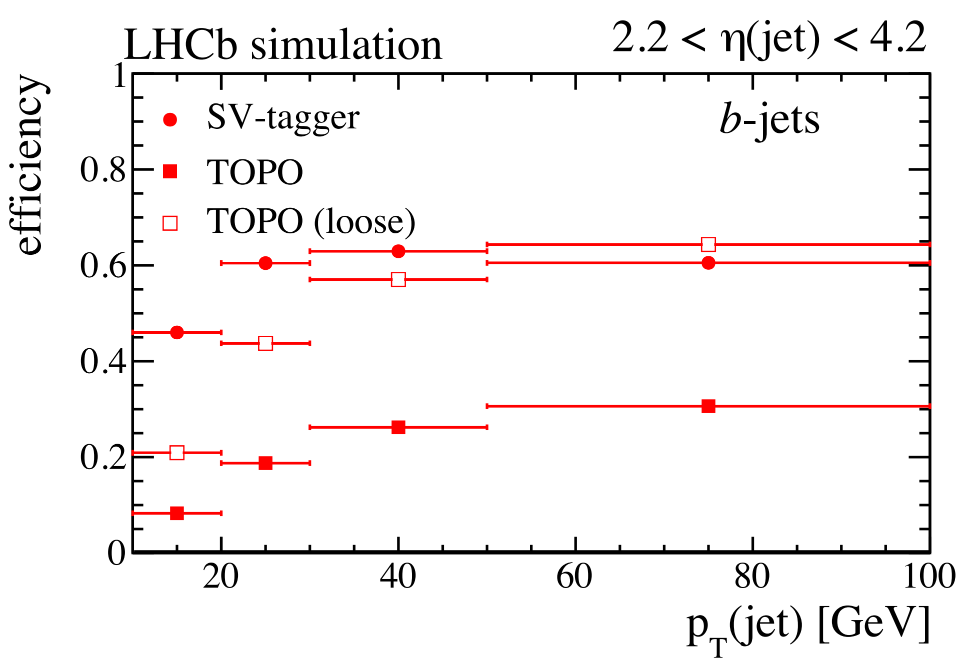
2. Optional Boosted Decision Trees (BDTs)

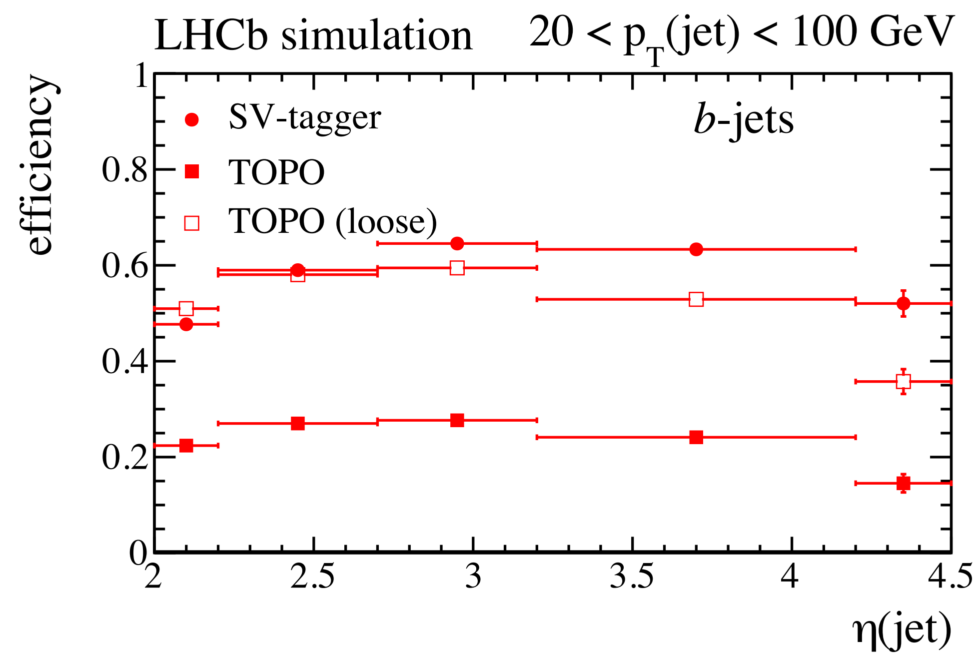
- Use input information about the SV in the jet to provide further flavor discrimination between beauty, charm, and light-parton jets



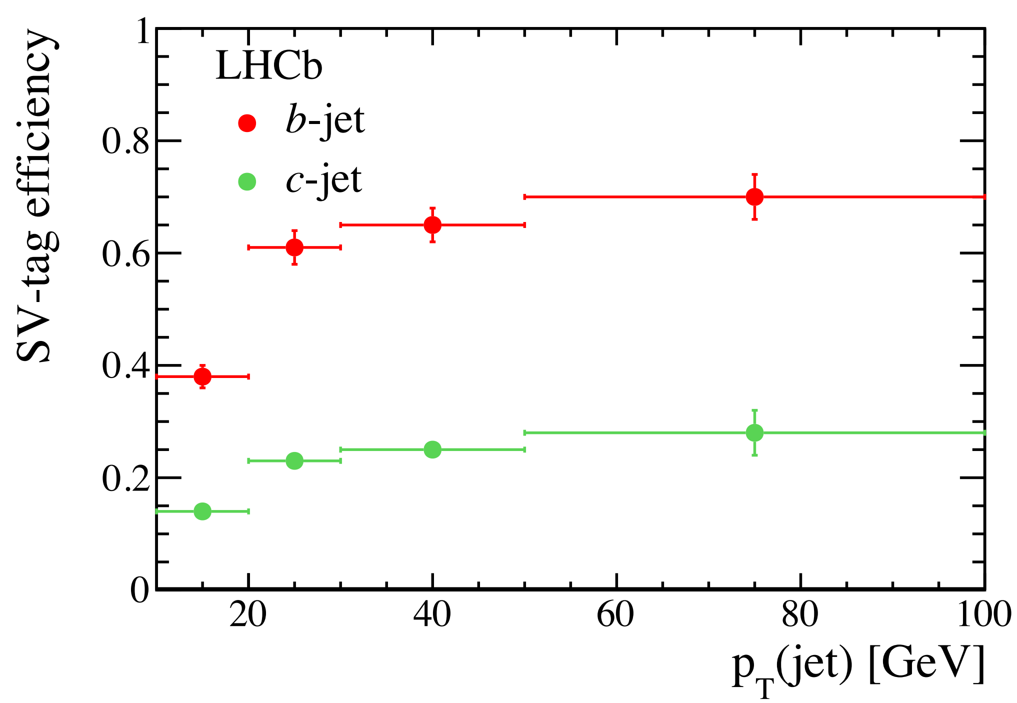
SV-tagging algorithm performance

- Comparison between SV-tagger and TOPO tagger (BDT for identification of b-hadron decays in the LHCb trigger):



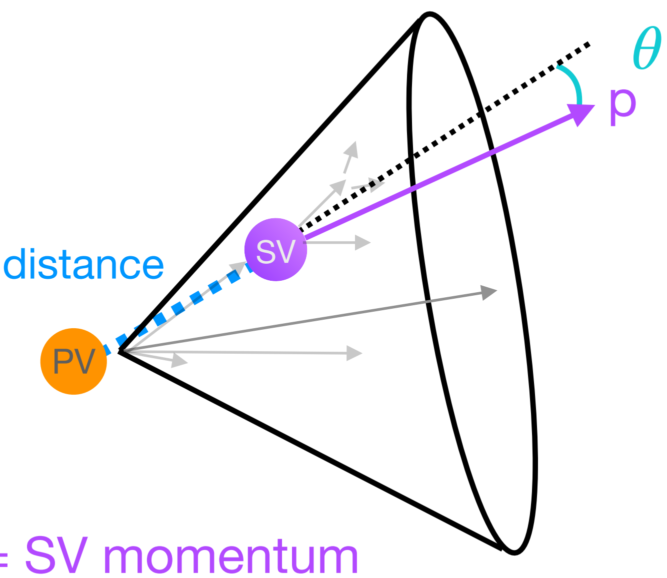


- For Run 1 data, $2.2 < \eta_{jet} < 4.2$, 20 < jet p_T < 100 GeV:
  - b-jet tagging efficiency of ~65%
  - c-jet tagging efficiency of ~25%
  - light jet mis-ID probability ~0.3%
  - Slight degradation in Run 2 data due to increased track multiplicity



Discriminating beauty from charm jets: BDT inputs

- SV flight distance χ^2
- Sum of the difference in χ^2 when performing a PV fit with and without each track in the SV
- Secondary vertex mass M
- Secondary vertex corrected mass M_{corr}: the minimum mass a long-lived hadron can have that is consistent with the SV flight direction

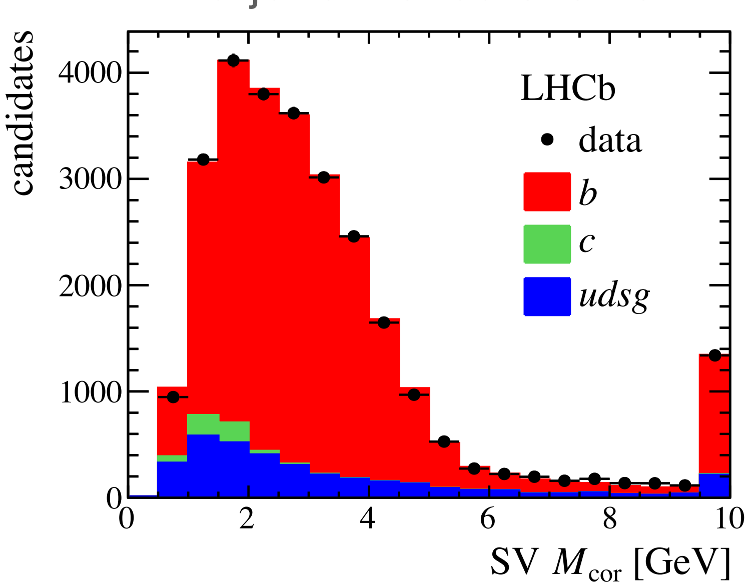


p = SV momentum

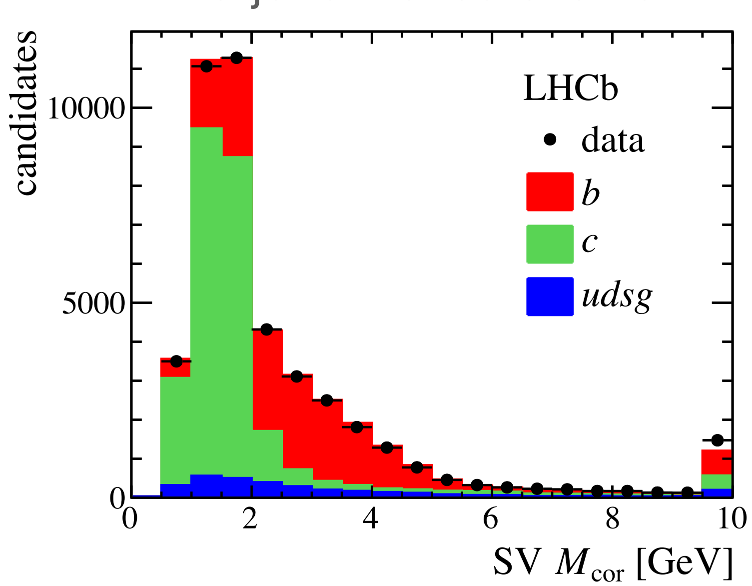
θ = angle between SV momentum and flight direction

$$M_{corr} \equiv \sqrt{M^2 + p^2 \sin^2\theta} + p \sin\theta$$

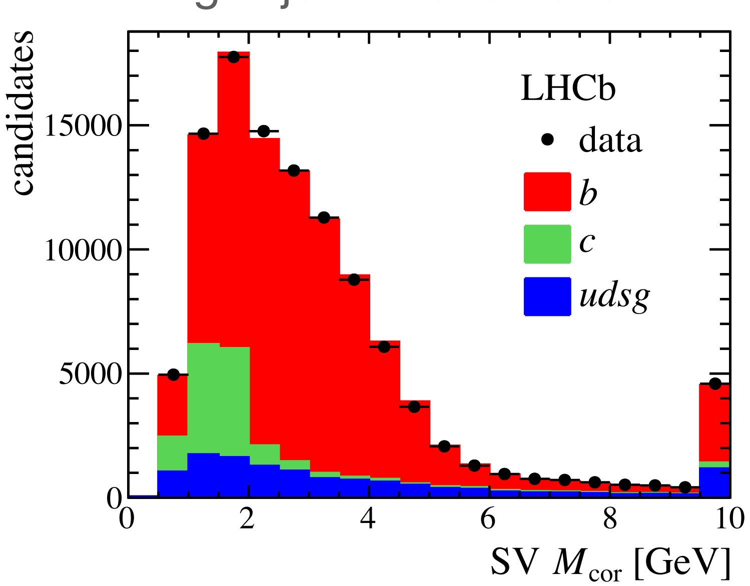
b-jet enriched events



c-jet enriched events



light-jet enriched events

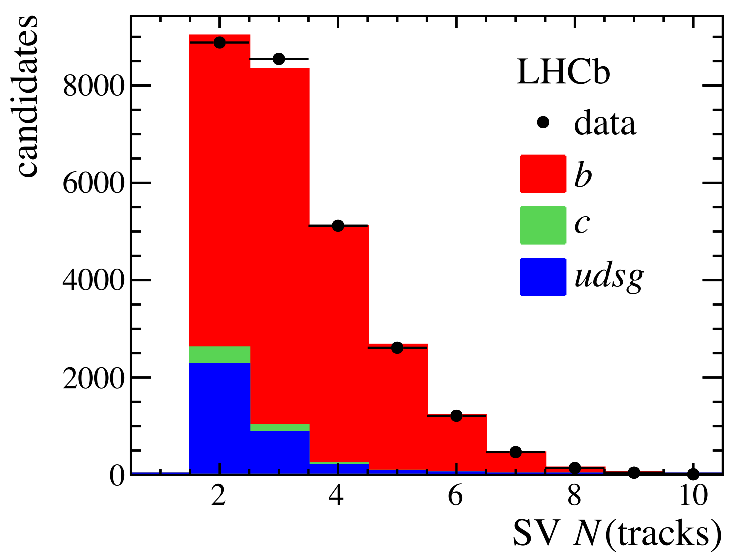


JINST 10, P06013 (2015)

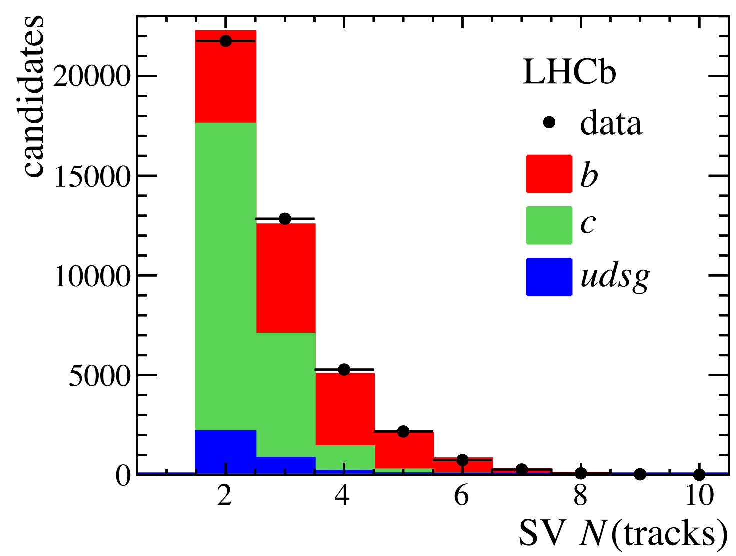
Discriminating beauty from charm jets: BDT inputs

- Number of tracks in the SV - excellent discriminator for b-jet tagging

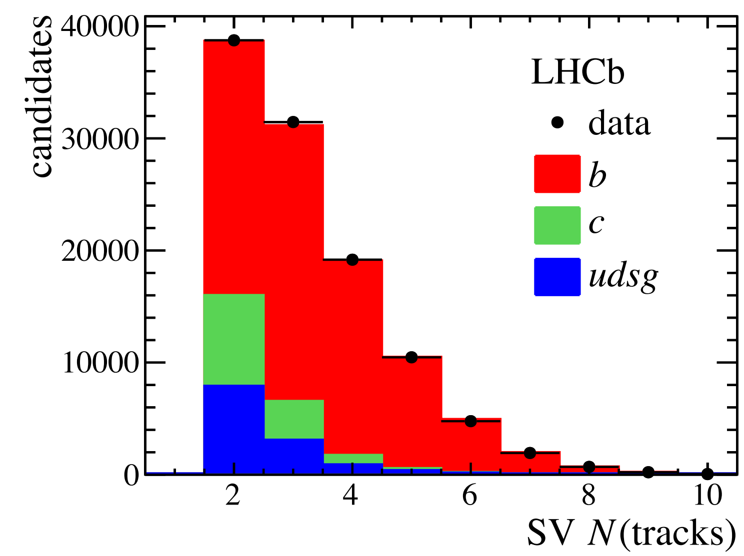
b-jet enriched events



c-jet enriched events



light-jet enriched events



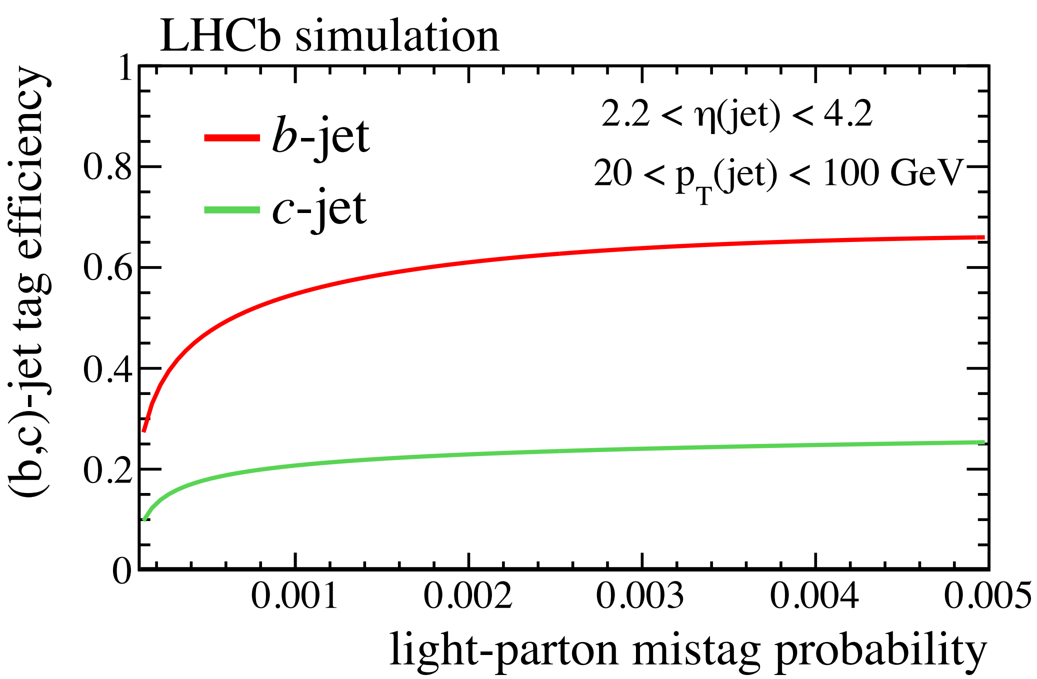
- Number of tracks in the SV with ΔR < 0.5 relative to the jet axis
- Net charge of the tracks that form the SV
- Transverse flight distance of the 2-track SV closest to the PV
- Fraction of the jet p_T carried by the SV: $p_T(SV)/p_T(jet)$
- ΔR between the SV flight direction and the jet

 JINST 10, P06013 (2015)

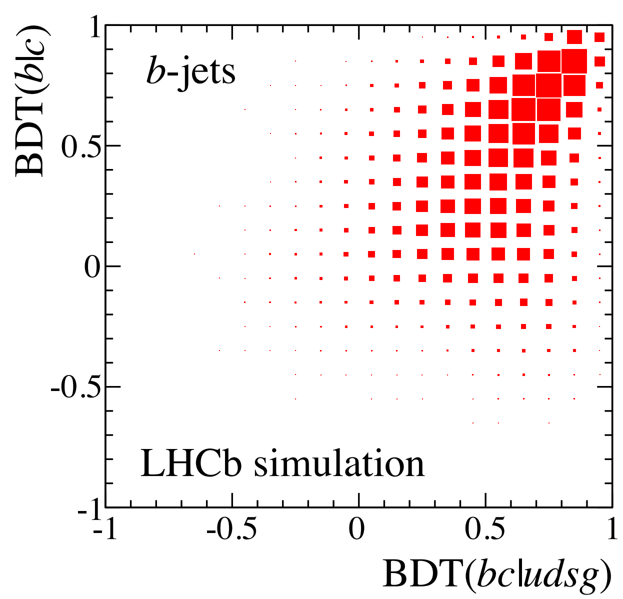
Boosted Decision Trees for jet flavor discrimination

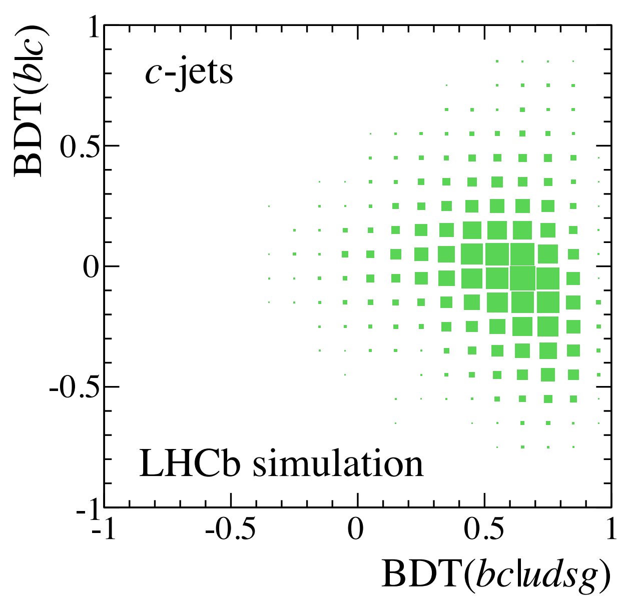
Two BDTs are used:

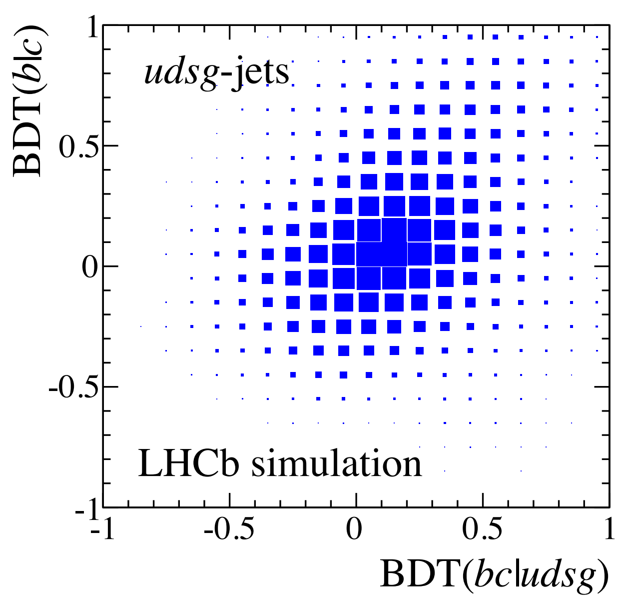
- BDT(bc|udsg) discriminates between heavy- and light-flavor jets
- BDT(b|c) discriminates between beauty and charm jets



2D BDT distributions clearly distinguish each jet type:





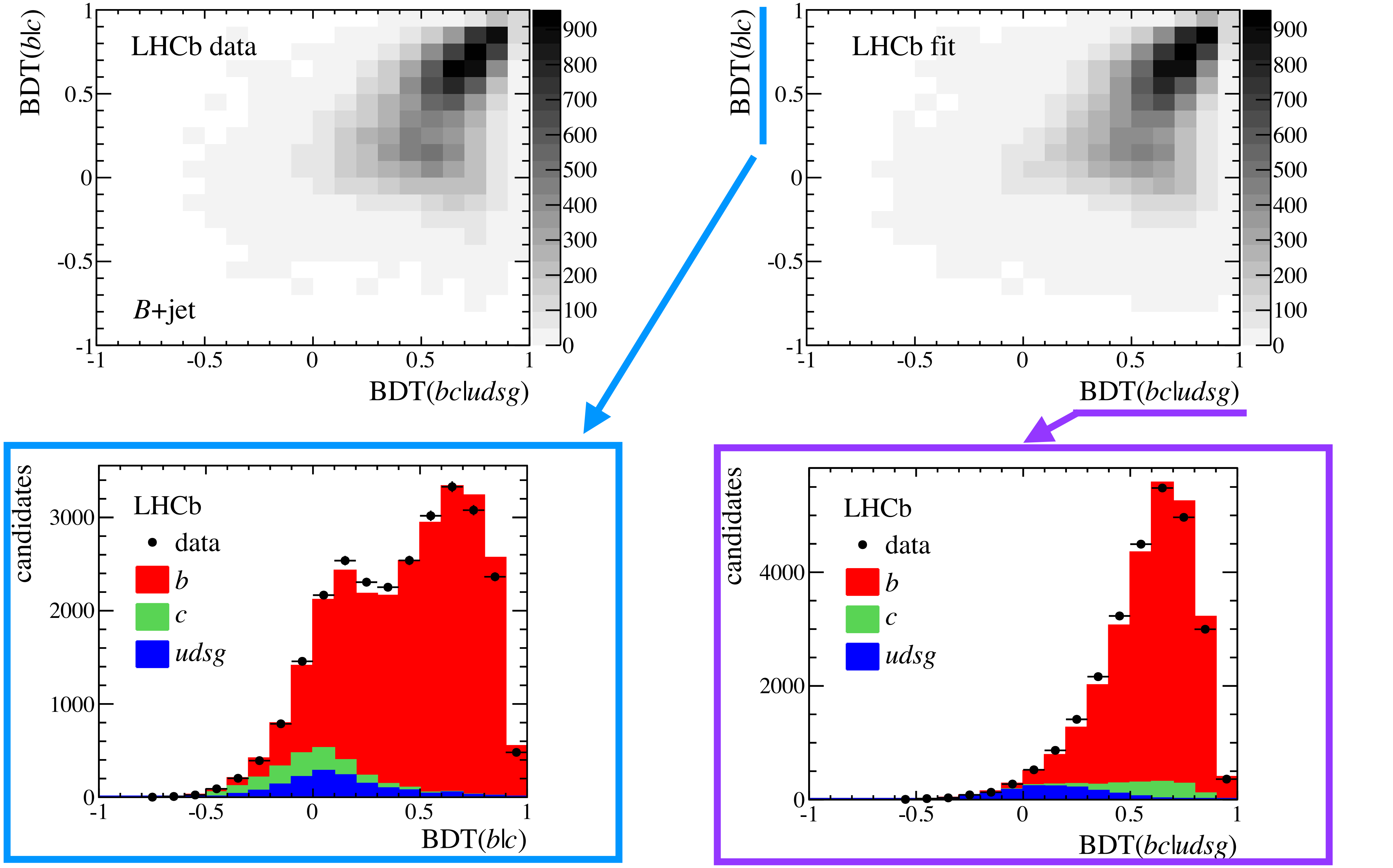


Use as templates for fit to 2D BDT distribution in data

JINST 10, P06013 (2015)

BDT template fits

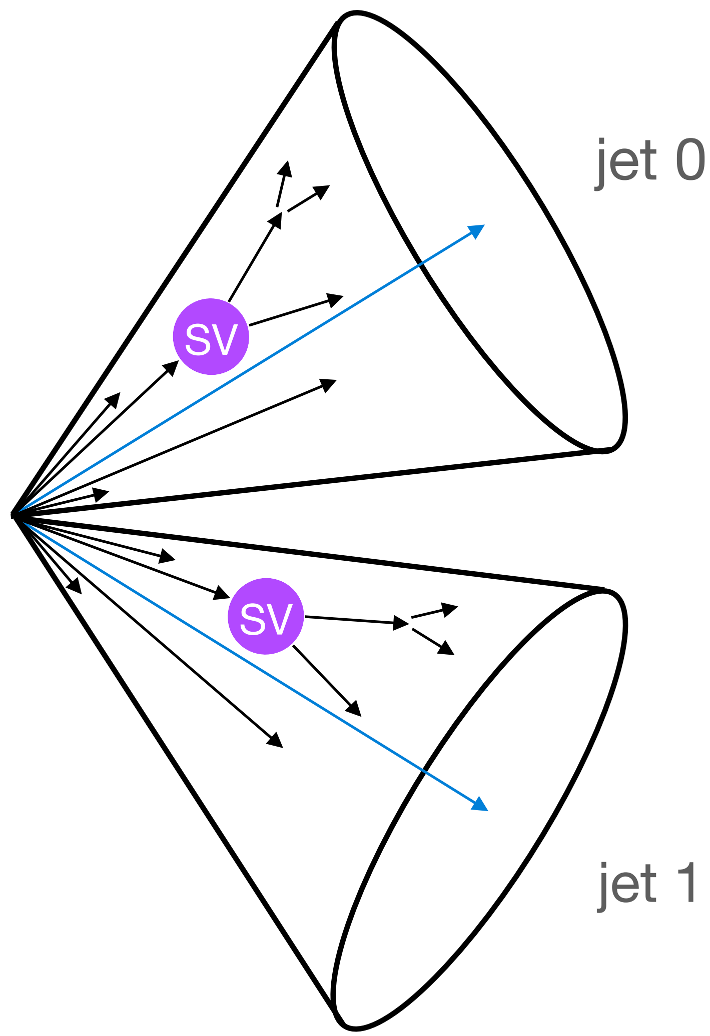
Template fit to 2D BDT distribution in data allows for extraction of flavor-tagged jet yields:



Heavy flavor dijet measurements

Forward $b\bar{b}$- and $c\bar{c}$-dijet cross sections in *pp* collisions at $\sqrt{s}$ =13 TeV

- Differential measurements allow for crucial tests of perturbative NLO QCD calculations
- Important background for BSM physics searches
- Measured at LHCb with 1.6 fb^{-1} 2016 dataset as a function of four variables:
  - Leading jet η
  - Leading jet p_T
  - Dijet invariant mass m_{jj}
  - Rapidity difference between jets, $\Delta y = (y_0 - y_1)/2$



- Both jets required to be SV-tagged and have p_T > 20 GeV and $2.2 < \eta < 4.2$, azimuthal angle between jets $|\Delta\phi| > 1.5$

Forward $b\bar{b}$- and $c\bar{c}$-dijet cross sections in *pp* collisions at $\sqrt{s} = 13$ TeV

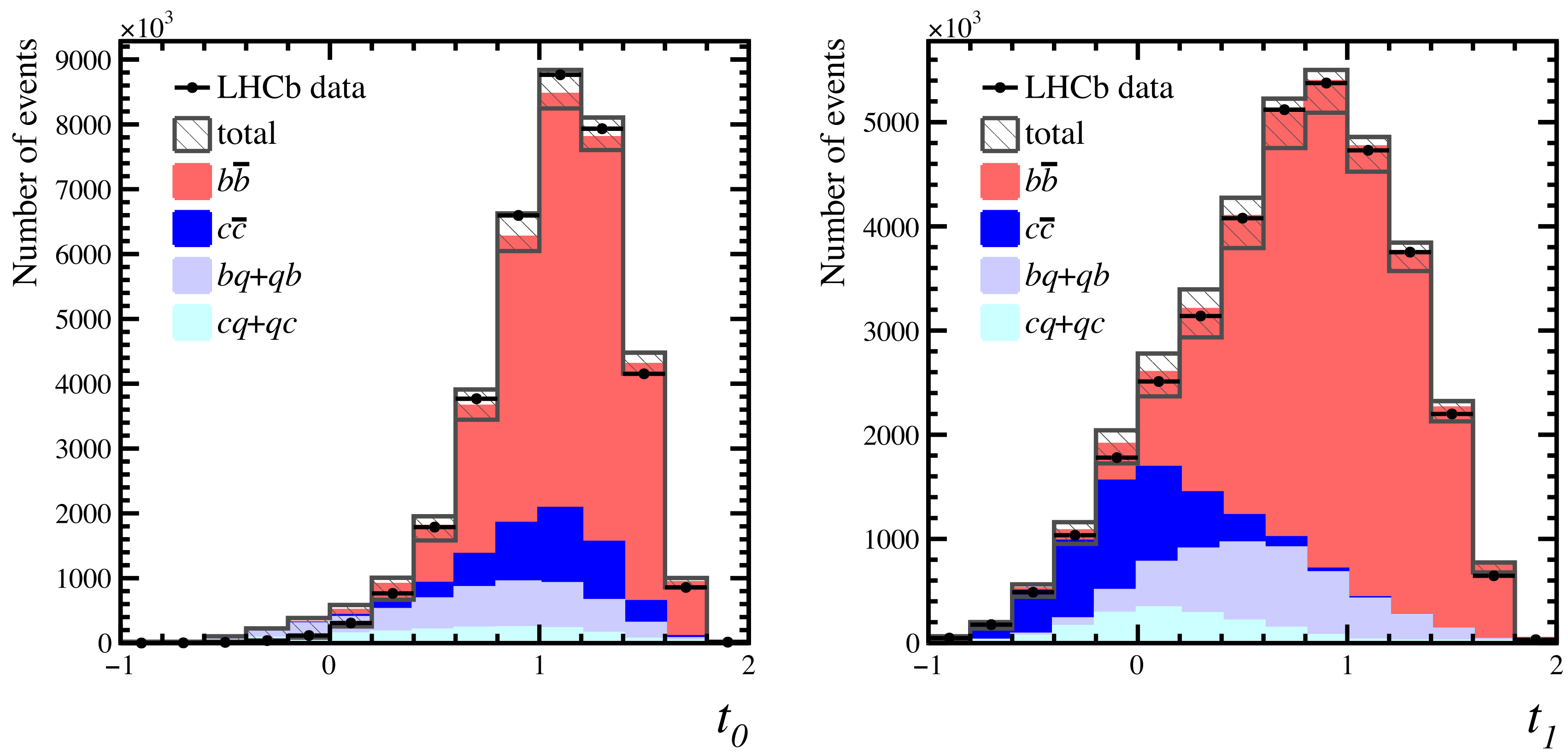
Flavor-tagged dijet yields:

- Four BDT scores needed to identify $b\bar{b}$ and $c\bar{c}$ events: BDT(bc|udsg) and BDT(b|c) for each jet
- Construct linear combinations of the BDT scores:

$$t_0 = \mathrm{BDT}_{bc|udsg}(\text{jet } 0) + \mathrm{BDT}_{bc|udsg}(\text{jet } 1)$$

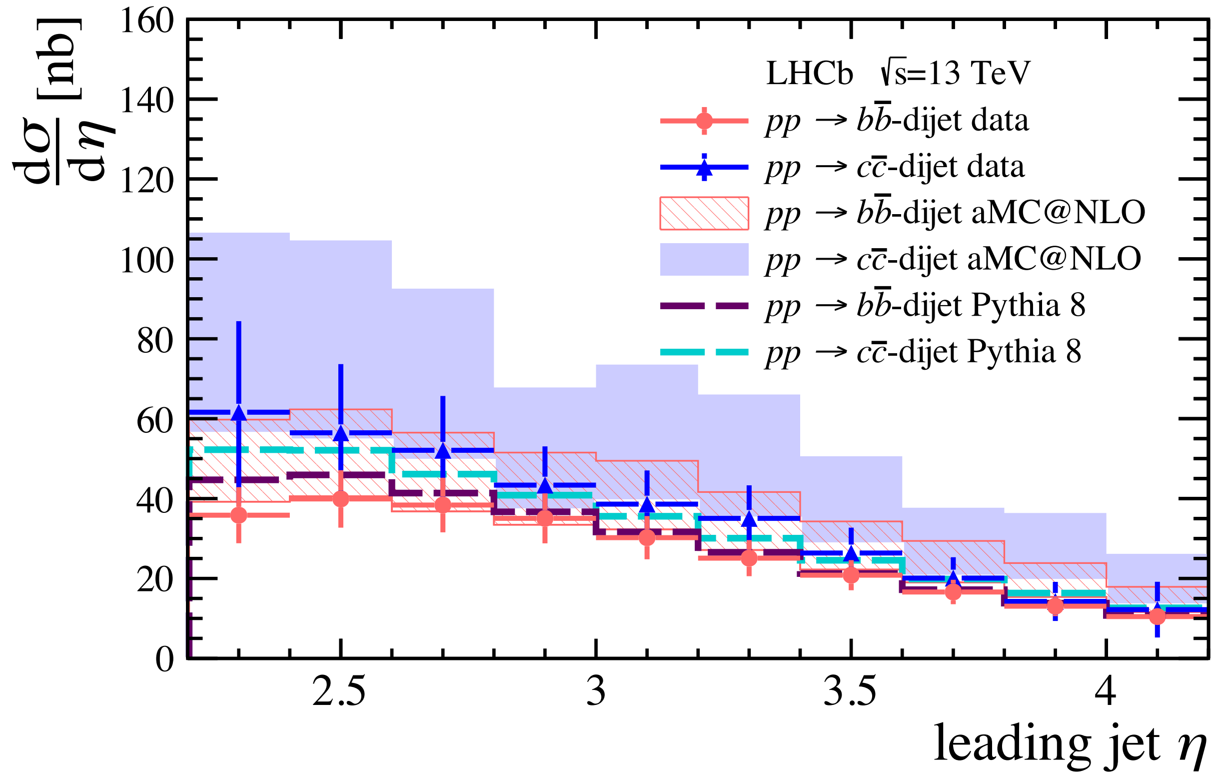
$$t_1 = \mathrm{BDT}_{b|c}(\text{jet } 0) + \mathrm{BDT}_{b|c}(\text{jet } 1)$$

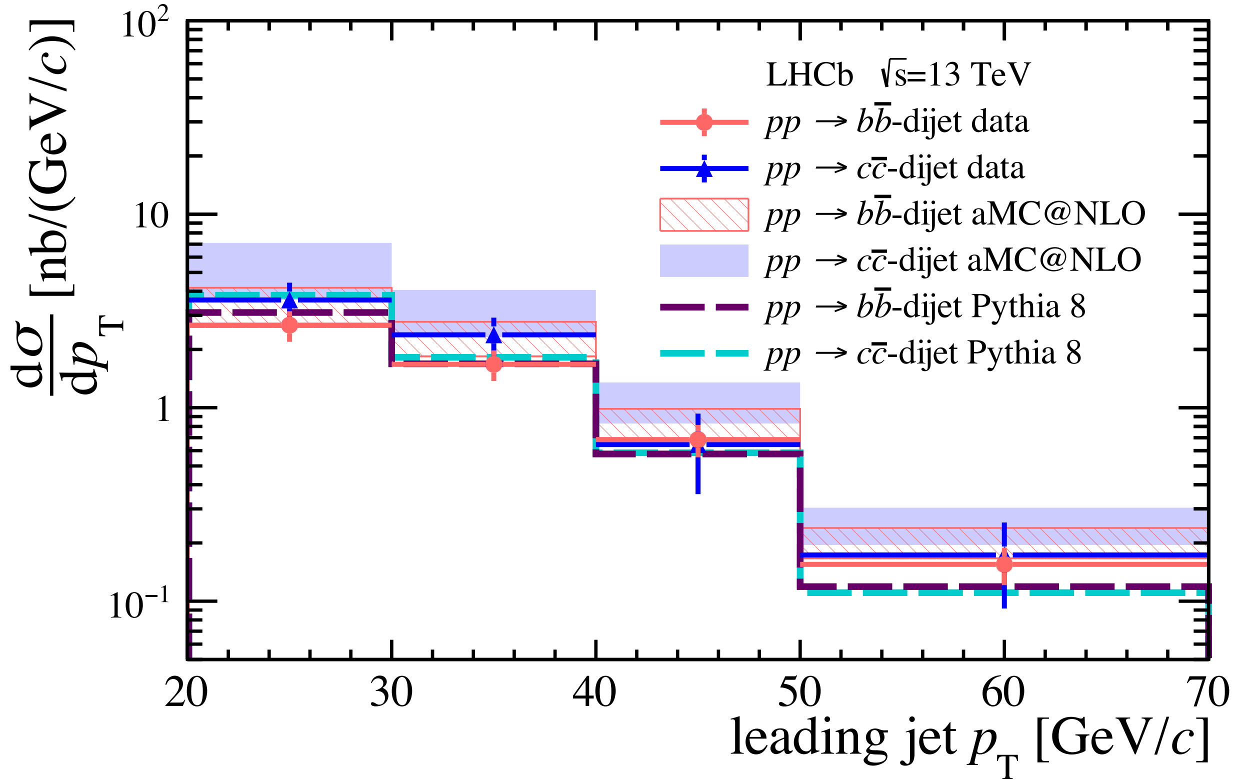
- Construct 2D (t_0, t_1) templates from simulation and fit to 2D (t_0, t_1) distribution in data
  - Fits performed in bins of jet p_T



Forward $b\bar{b}$- and $c\bar{c}$-dijet cross sections in *pp* collisions at $\sqrt{s}$ =13 TeV

Differential cross sections: Leading jet η and p_T

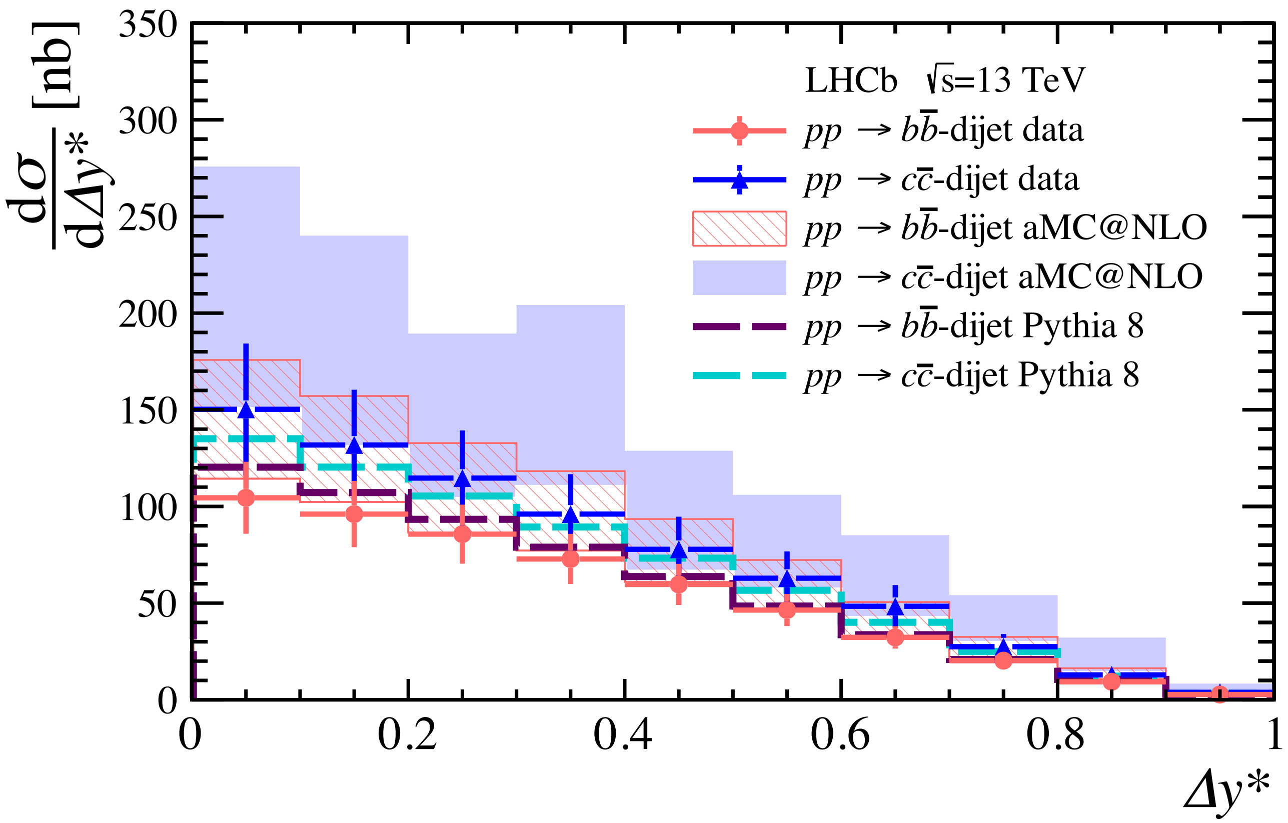


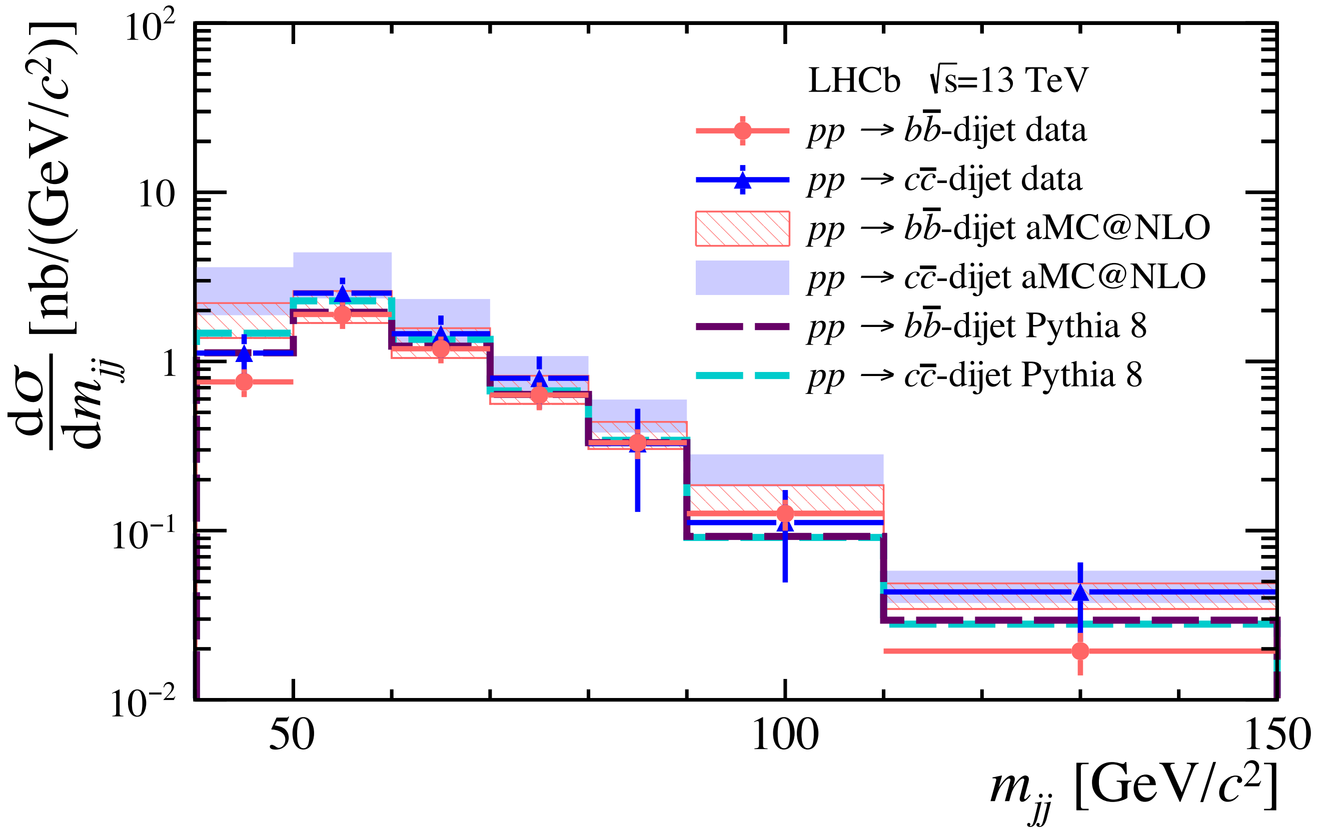


- **First $c\bar{c}$-dijet differential cross section measurement at a hadron collider!**
- LO pQCD predictions obtained with Pythia 8
- NLO pQCD predictions obtained with Madgraph5 aMC@NLO for matrix element computation + Pythia for parton showers

Forward $b\bar{b}$- and $c\bar{c}$-dijet cross sections in *pp* collisions at $\sqrt{s}$ =13 TeV

Differential cross sections: Δy* and m_{jj}

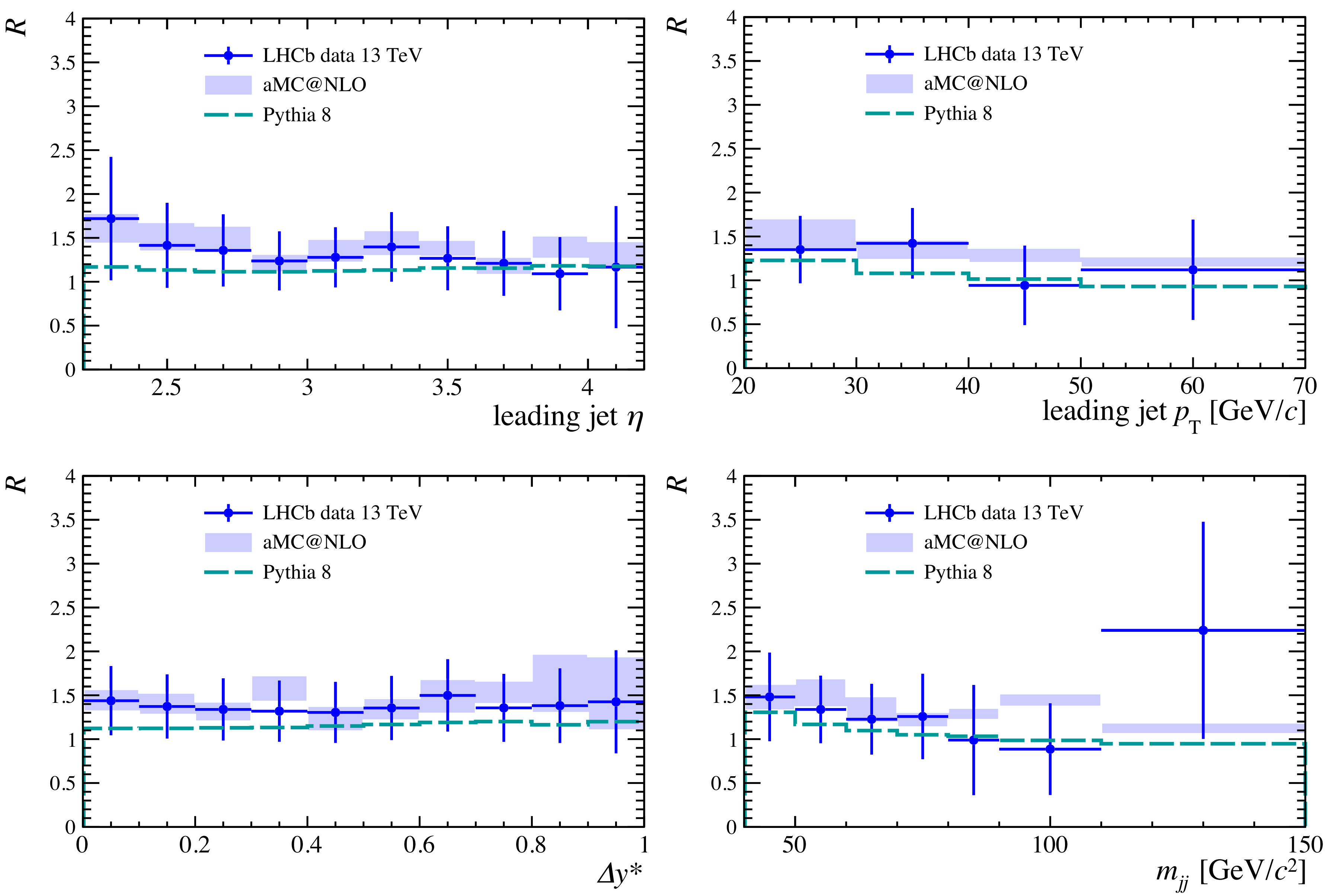




- Measurements are generally below the NLO pQCD theory predictions
- Theory uncertainties are dominated by the renormalization and factorization scale uncertainties at low m_{jj} and leading jet p_T

Forward $b\bar{b}$- and $c\bar{c}$-dijet cross sections in *pp* collisions at $\sqrt{s}$ =13 TeV

$c\bar{c}$ to $b\bar{b}$ cross section ratios:

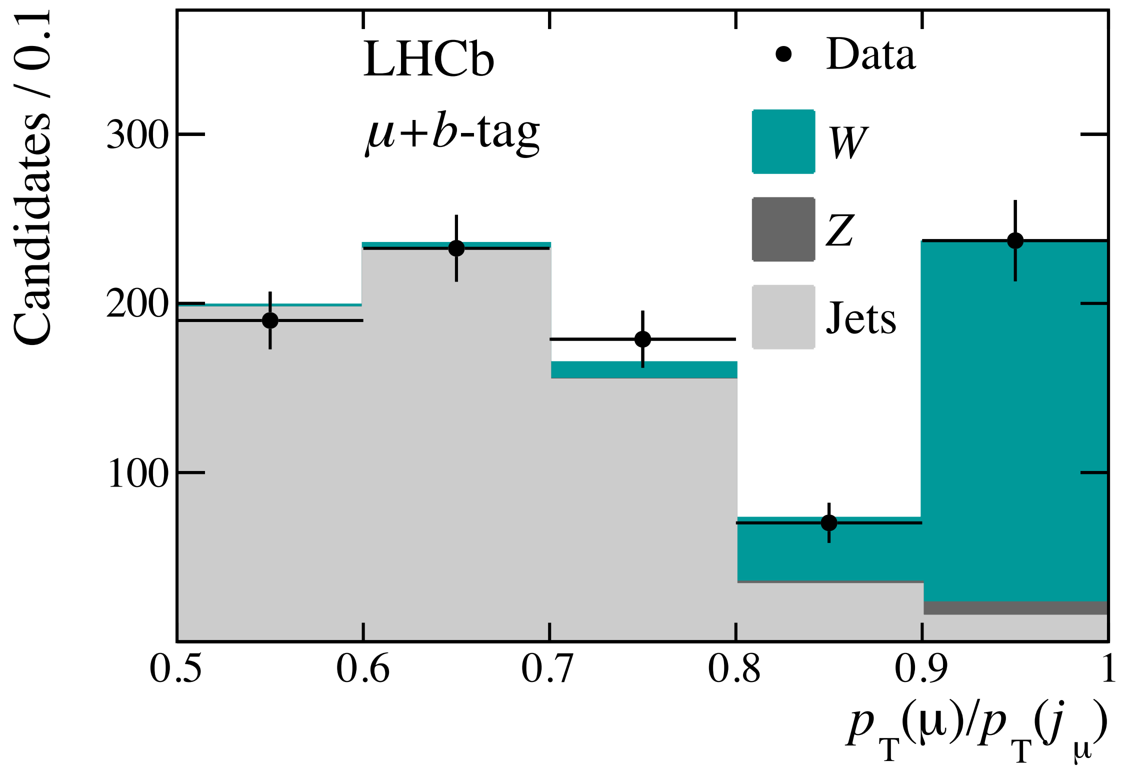


- Uncertainty on ISR and FSR contributes 2.4% systematic uncertainty to measurement
  - Correlated between kinematic bins
- Measured ratio R = 1.37 ± 0.27
  - Much lower than expected inclusive $c\bar{c}/b\bar{b}$ ratio due to jet p_T cuts in fiducial region

Top production in the forward region

Top production in *pp* collisions at $\sqrt{s}$ =7 and 8 TeV

- Enhancement of $t\bar{t}$ production in the forward region from *qq* and *qg* scattering can result in larger charge asymmetries, which could be sensitive to new physics
  - Forward top production dominated by $t\bar{t}$ pair production (~75%)
- First measurement of top production at LHCb performed with 3 fb^{-1} of data in the μ + *b*-jet final state
  - 50 < p_T(*b*-jet) < 100 GeV
  - p_T(μ + *b*-jet) > 20 GeV, ΔR(μ, *b*-jet) > 0.5
  - High-p_T muon (p_T > 25 GeV) clustered into a jet, j_μ to distinguish dijet and W+jet events
  - No differentiation between single top and $t\bar{t}$ production



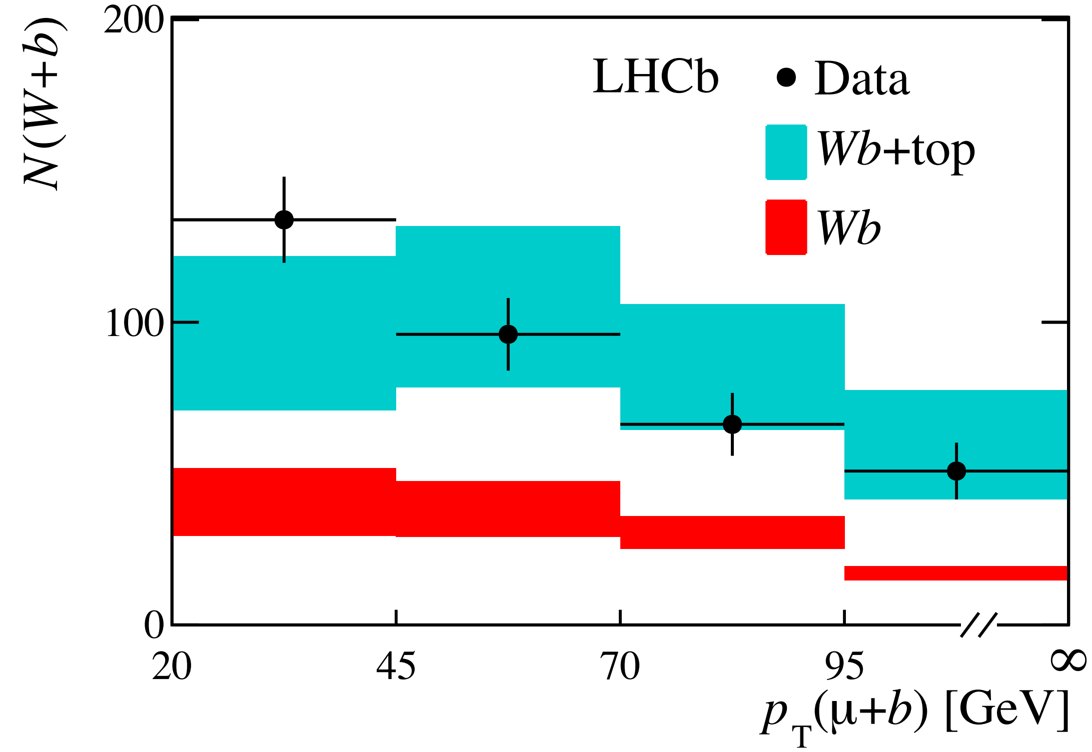
Top production in *pp* collisions at $\sqrt{s}$ =7 and 8 TeV

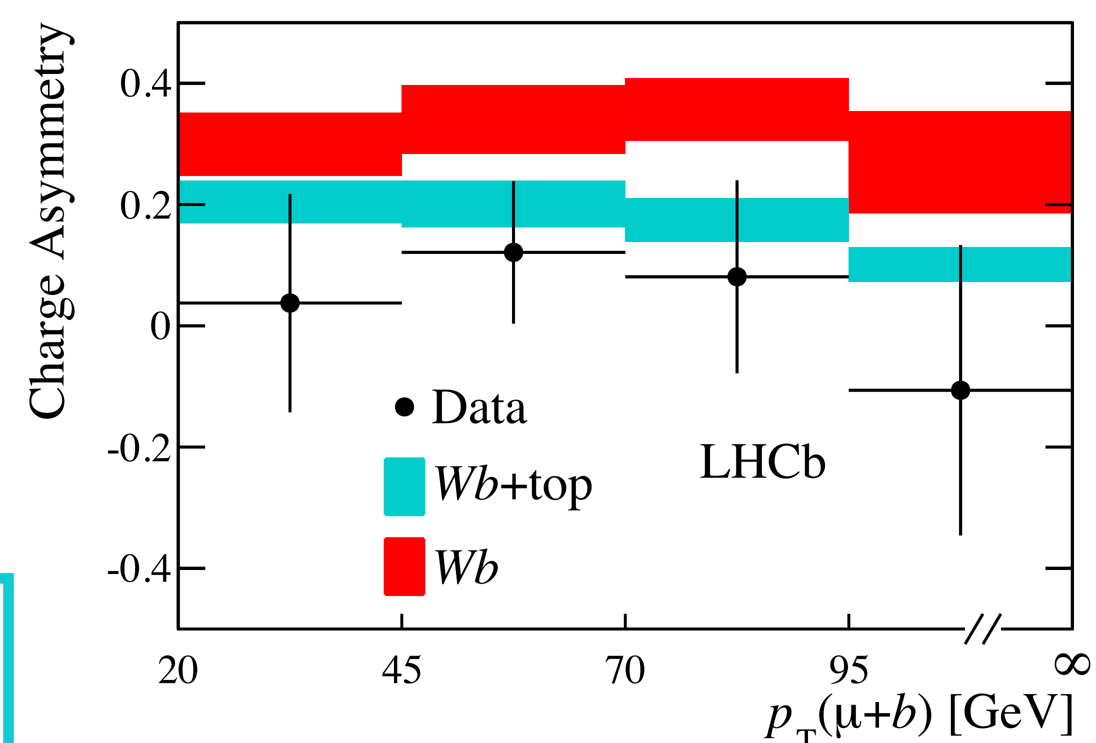
- Top yield determined as the excess of the observed yield relative to the direct W+*b* prediction
- W+b predictions determined from σ(W+*b*)/σ(W+jet), which has smaller uncertainties than σ(W+*b*) alone:

$$N(W+b)_{theory} = \left(\frac{\sigma(W+b)}{\sigma(W+jet)}\right)_{theory} \times N(W+jet)_{measured} \times \varepsilon_{b\text{-}tag}$$

- Theory calculations are performed at NLO with MCFM and CT10 PDFs
  - Uncertainties 100% correlated across $p_T(\mu+b)$ bins

$$\sigma(\text{top})[7\,\text{TeV}] = 239 \pm 53\,(\text{stat}) \pm 33\,(\text{syst}) \pm 24\,(\text{theory})\,\text{fb}$$
$$\sigma(\text{top})[8\,\text{TeV}] = 289 \pm 43\,(\text{stat}) \pm 40\,(\text{syst}) \pm 29\,(\text{theory})\,\text{fb}$$

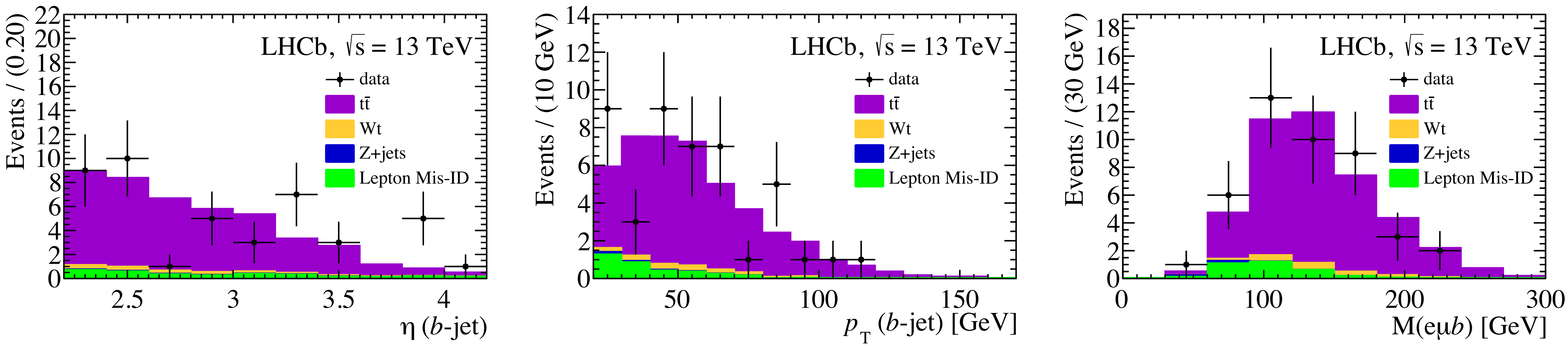




PRL 115, 112001 (2015)

Forward top pair production in *pp* collisions at √s =13 TeV

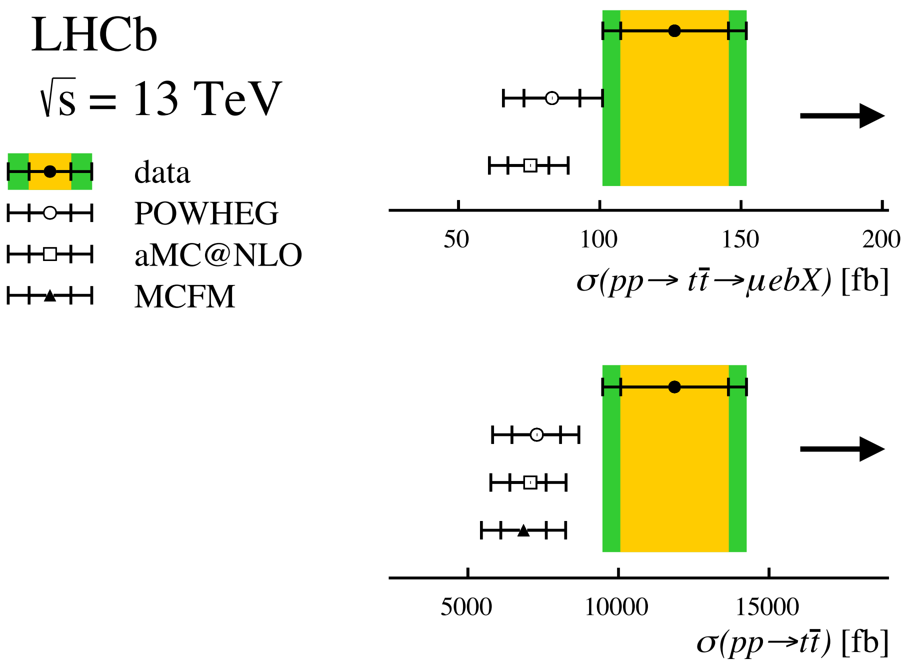
- $t\bar{t}$ production measured in the *μeb* final state with 1.93 fb^{-1} 2015+2016 dataset
  - Lepton requirements: $p_T(e/\mu) > 20$ GeV, $2.2 < \eta(e/\mu) < 4.2$, $\Delta R\ (e,\mu) > 0.1$
  - *b*-jet requirements: SV-tagged, $p_T(\text{jet}) > 20$ GeV, $2.2 < \eta(\text{jet}) < 4.2$, $\Delta R\ (e/\mu,\ \text{jet}) > 0.5$
- 44 events selected, 5.6 expected from background



Forward top pair production in *pp* collisions at $\sqrt{s}$ =13 TeV

- Measurement systematics impacted by theory uncertainties:
  - W+top background determined with POWHEG in the diagram removal scheme
  - Reconstruction and identification efficiencies determined after reweighting distributions from Pythia 8 in p_T and η to match NLO distributions from aMC@NLO

Source	%
trigger	2.0
muon reconstruction	1.1
electron reconstruction	2.8
muon identification	0.8
electron identification	1.3
jet reconstruction	1.6
event selection	4.0
jet tagging	10.0
background	5.1
resolution factor	0.5
total	12.7



Measurement in fiducial region of μeb final state

Extrapolation of measurement to full phase space using predictions from aMC@NLO

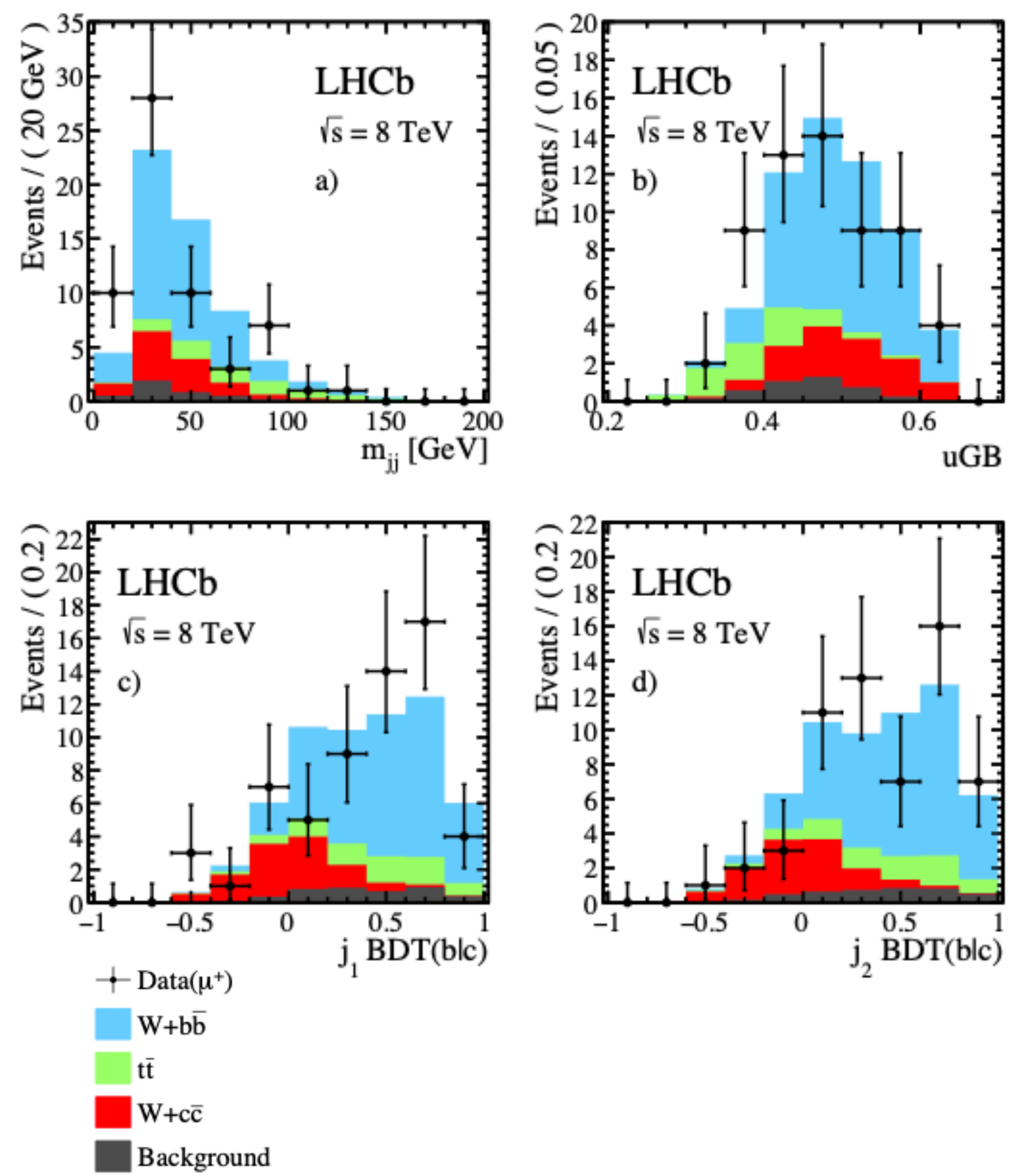
- Contributes another 1.5% uncertainty

(W/Z) + HF jet production

$t\bar{t}$, W+$b\bar{b}$ and W+$c\bar{c}$ cross sections in *pp* collisions at $\sqrt{s}$ = 8 TeV

- Events required to have a high-p_T lepton (p_T > 20 GeV) and two SV-tagged jets (jet p_T> 12.5 GeV)

- Simultaneous fit to four distributions performed to extract $t\bar{t}$, W^++$b\bar{b}$, W^-+$b\bar{b}$, W^++$c\bar{c}$, and W^-+$c\bar{c}$ yields:
  - Invariant mass of the two HF jets, m_{jj}
  - Response of multivariate classifier trained to distinguish $t\bar{t}$ and W+$b\bar{b}$ events, uGB
  - BDT(b|c) response for each HF jet

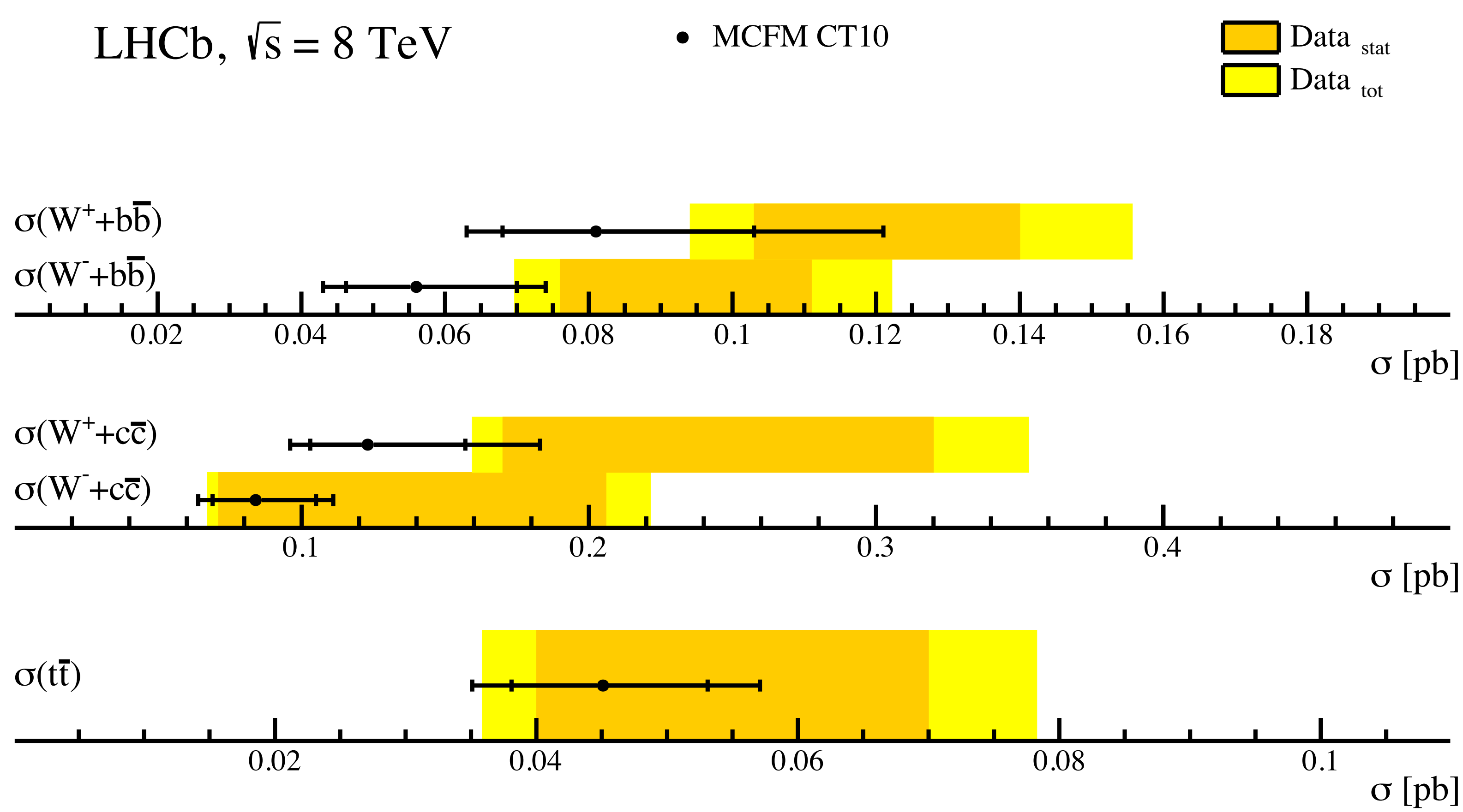


PLB 767, 110 (2017)

$t\bar{t}$, $W+b\bar{b}$ and $W+c\bar{c}$ cross sections in *pp* collisions at $\sqrt{s}$ = 8 TeV

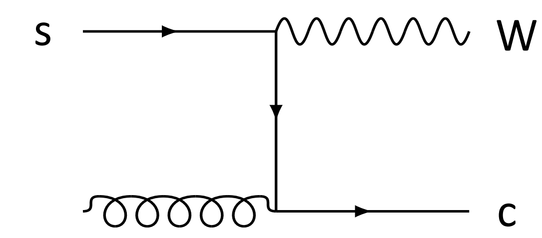
- Background process yields ($Z+b\bar{b}$, $Z+c\bar{c}$, W+light jets, single top, ZZ, WZ) are fixed from NLO cross sections and allowed to vary within theoretical uncertainties, contribute 3-10% relative uncertainty on signal yields
  - Correlated between $\mu^{\pm}$, $e^{\pm}$ data samples

- Measured cross sections compared to NLO theory calculations using MCFM and the CT10 PDFs
  - Inner bars on theory predictions are scale uncertainties

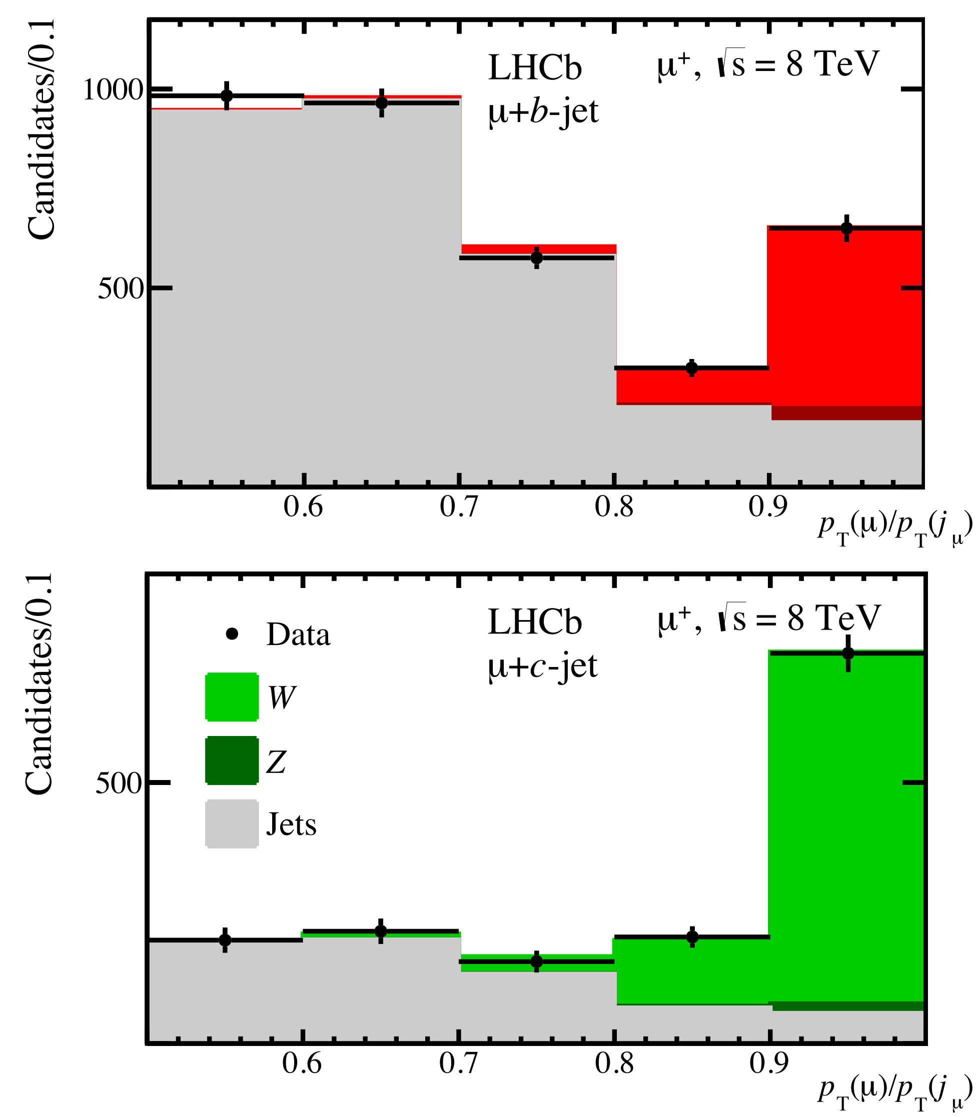


W+b and W+c production at $\sqrt{s}$ = 7 and 8 TeV

- W+c production is sensitive to the strange quark parton distribution function at leading order
  - In the forward region, can constrain the strange PDF at high and low-x

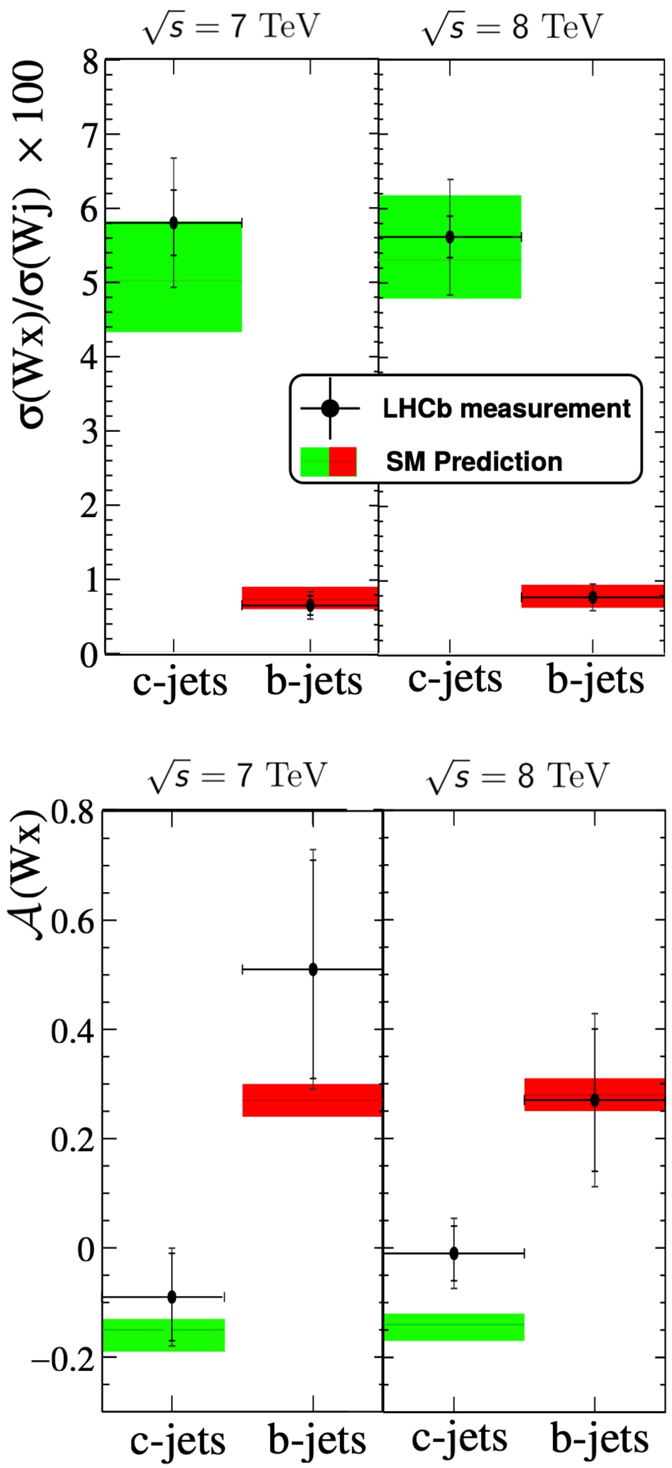


- W+b production is sensitive to the probability of gluon splitting into $b\bar{b}$ pairs (four-flavor scheme) and the intrinsic b-quark content of the proton (five-flavor scheme)
- Events required to have a high-p_T muon (p_T > 20 GeV) and a well-separated jet (p_T > 20 GeV, ΔR (μ, jet) > 0.5)
  - b and c-tagged yields determined from template fit to 2D BDT distribution



 PRD 92, 052001 (2015)

W+b and W+c production at √s = 7 and 8 TeV



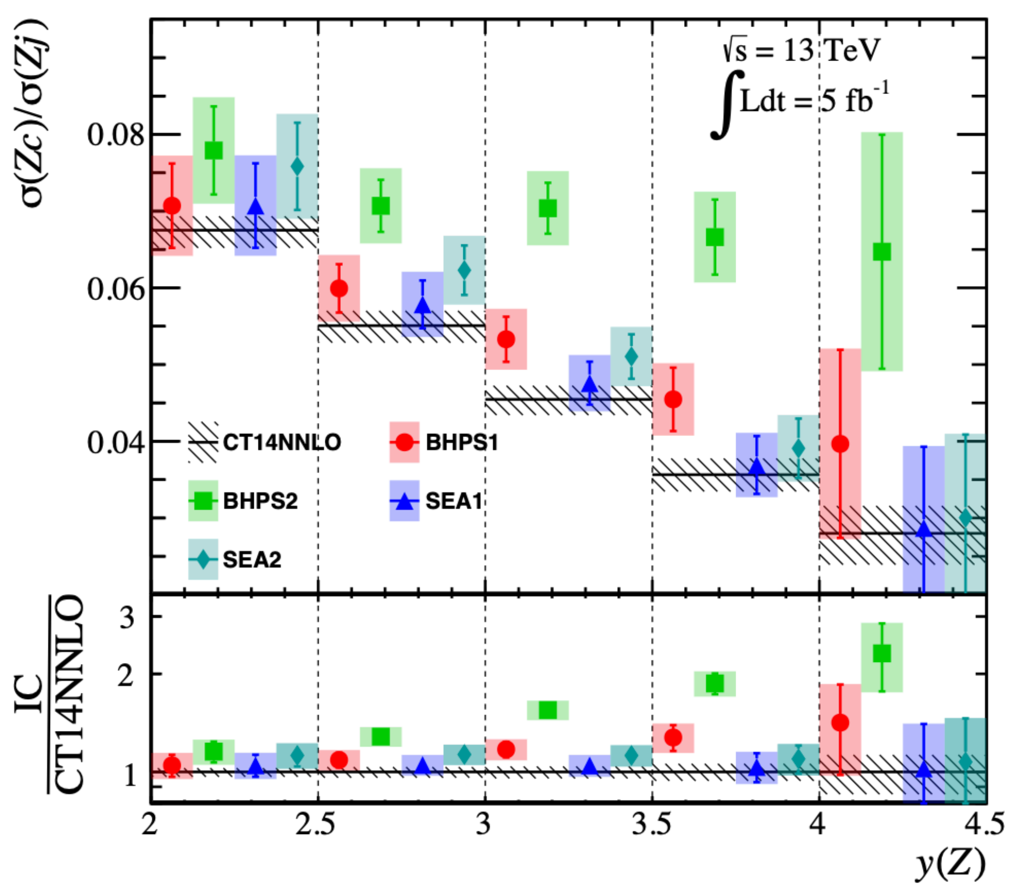
- SM predictions are calculated at NLO using MCFM and the CT10 PDFs
- σ(W+b)/σ(W+jet) and A(W+b) are consistent with MCFM calculations in the four-flavor scheme
  - Precision in data is not sufficient to completely rule out intrinsic b-quark content in the proton
- Discrepancy of ~2σ observed between measured W+c charge asymmetry and SM prediction.
  - CT10 PDFs assume symmetric s and s̄ quark PDFs
  - Discrepancy could indicate charge asymmetry between s and s̄ in proton, or a larger than expected scattering off of strange quarks

PRD 92, 052001 (2015)

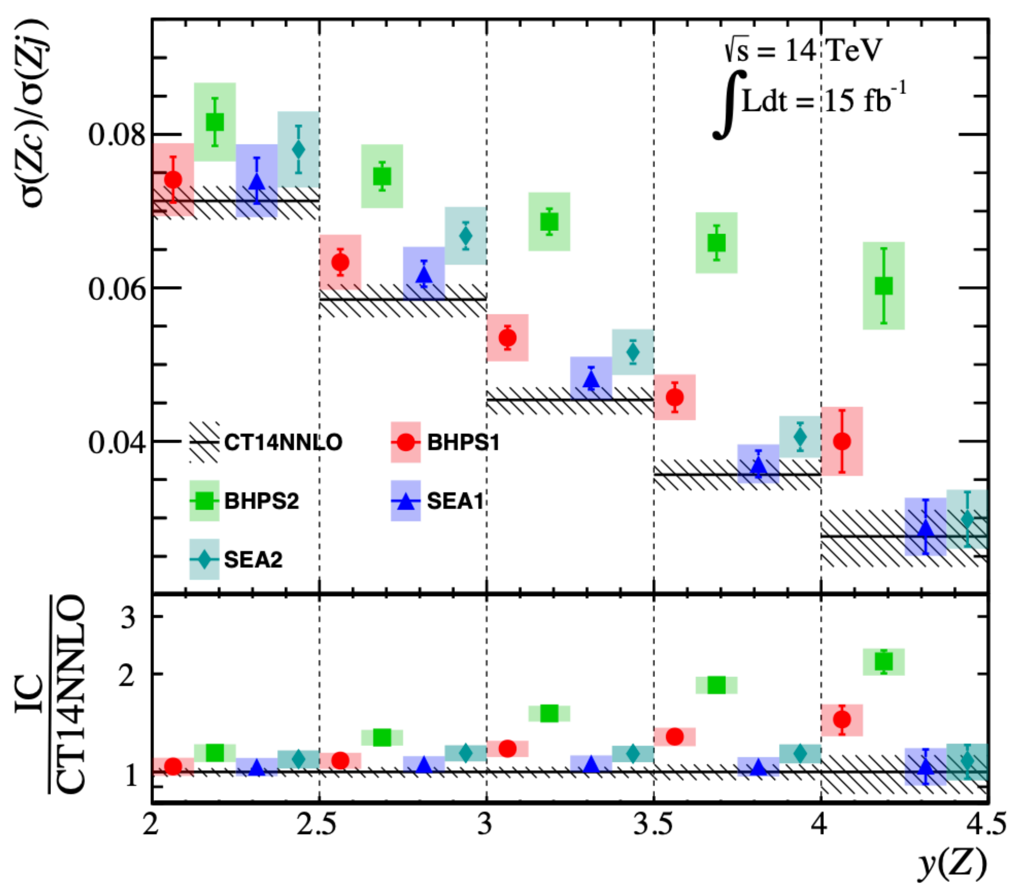
Probing intrinsic charm with Z+c

- Measurements of σ(Z+c)/σ(Z+jet) at LHCb are sensitive to the intrinsic charm quark content in the proton
- Run 2 statistics sufficient to detect existence of valence-like intrinsic charm with $\langle x \rangle \gtrsim 1\%$
  - Measurement in progress!

Run 2 projections:



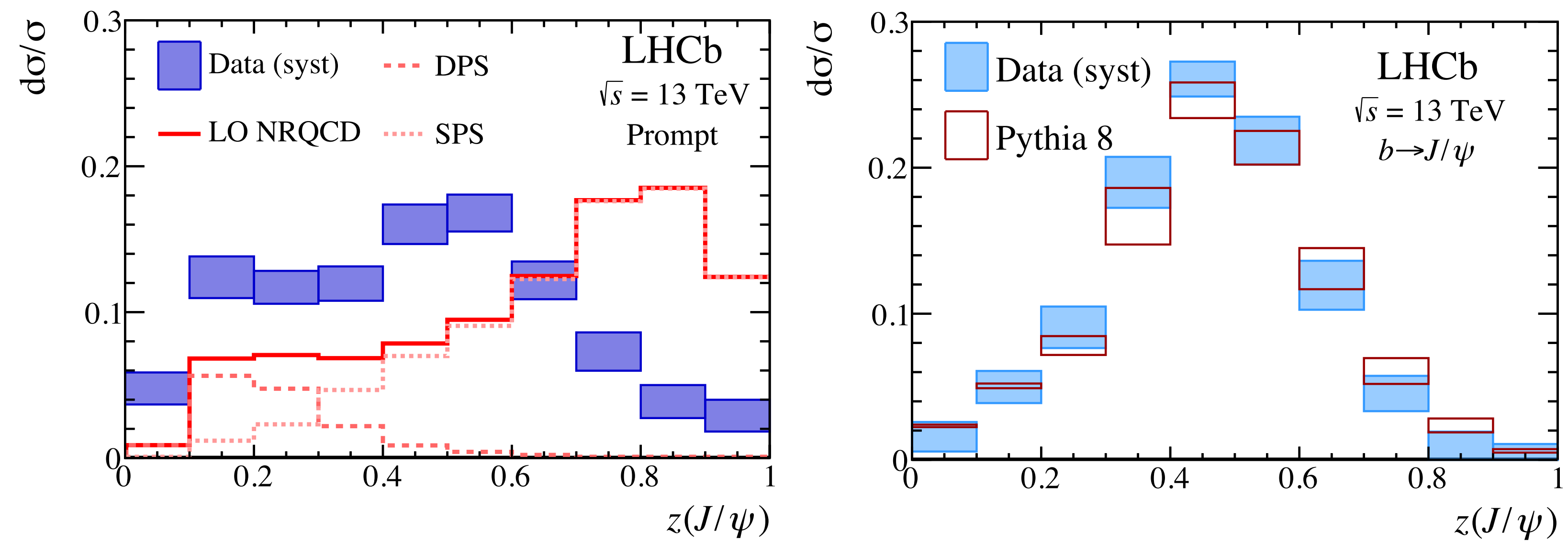
Run 3 projections:



Boettcher, Ilten, and Williams
PRD 93 , 074008 (2016)

Heavy flavor jet substructure measurements

PRL 118, 192001 (2017)



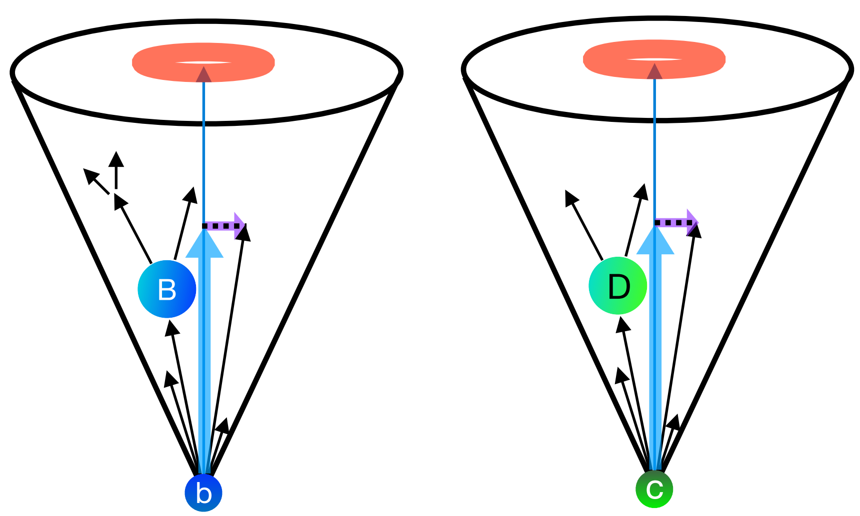
Ongoing measurements:

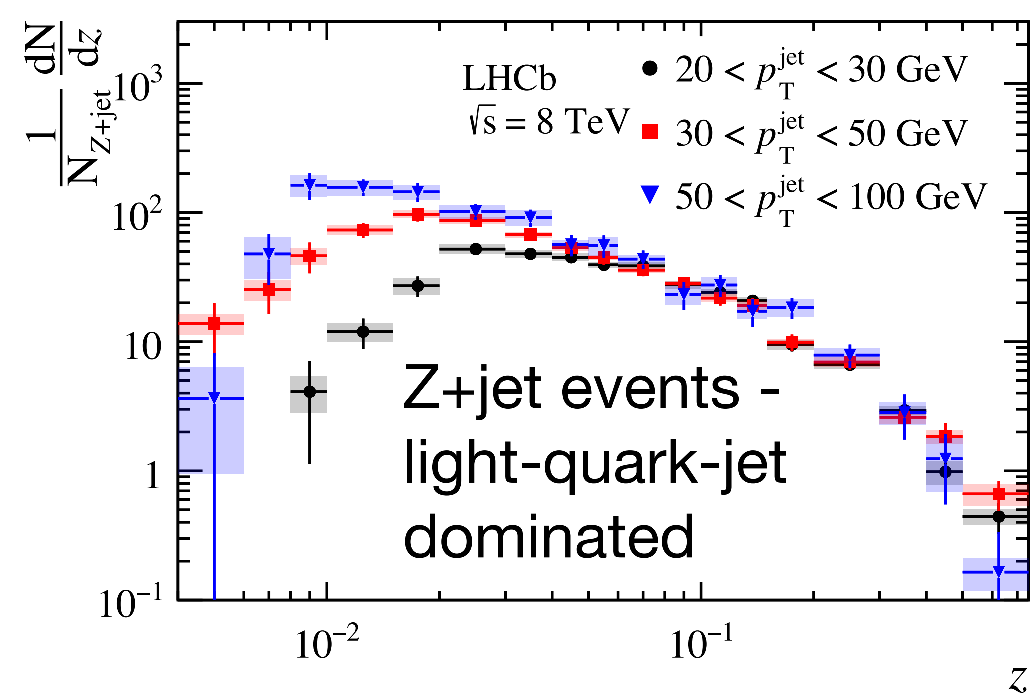
- Heavy quarkonia production in jets
  - J/ψ, Υ
- Hadronization distributions in b- and c-tagged jets

$$z = \frac{\mathbf{p_{jet}} \cdot \mathbf{p_{hadron}}}{|\mathbf{p_{jet}}|^2}$$

$$j_T = \frac{|\mathbf{p_{jet}} \times \mathbf{p_{hadron}}|}{|\mathbf{p_{jet}}|}$$

$$r = \sqrt{(\phi_{jet} - \phi_{hadron})^2 + (y_{jet} - y_{hadron})^2}$$





PRL 123, 232001 (2019)

Summary

- LHCb has performed several measurements of heavy flavor dijet production, top production, and (W/Z)+jet production
  - SV-tagging of heavy flavor jets is the dominant systematic uncertainty
- Measurements of W+c and Z+c in the forward region constrain the strange and intrinsic charm PDFs, respectively, at high and low x
- With increasing interest in studies of heavy flavor jet substructure and fragmentation:

  For inclusive b- and c-jet tagging: How can we use measurements of heavy flavor jet substructure to improve jet tagging techniques?

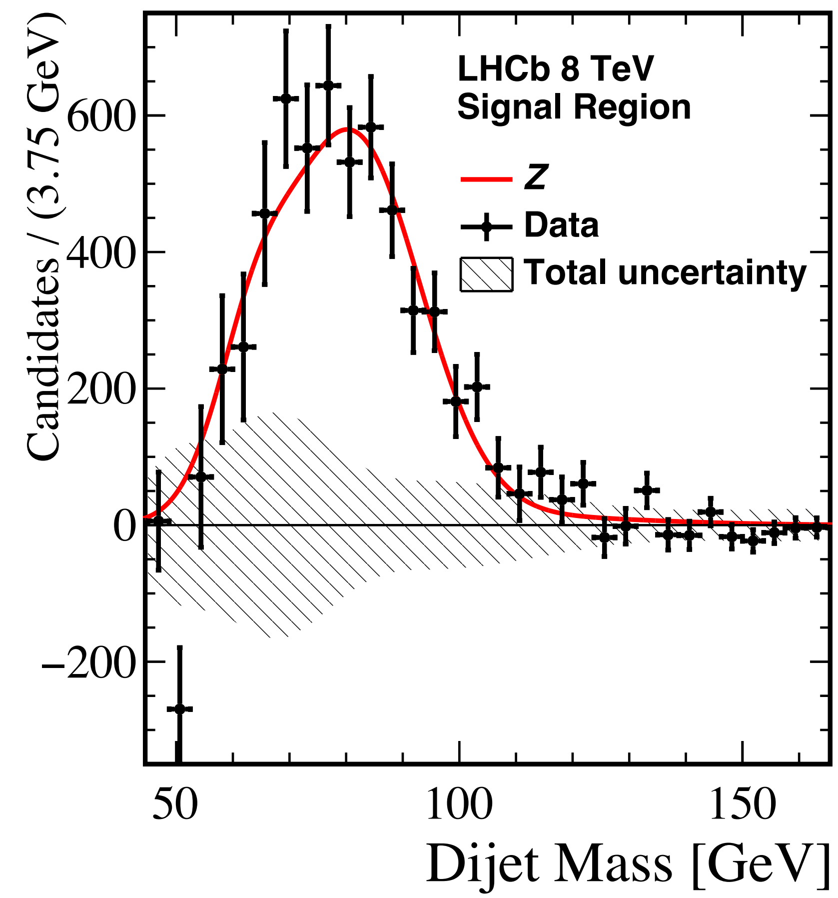
  For heavy flavor jet substructure measurements:

  - How can we check/correct for biases introduced by ML jet tagging techniques, or develop tagging techniques independent of jet substructure variables?
  - Advantages/disadvantages of using SV-tagged jets vs jets tagged with fully reconstructed heavy flavor hadrons?

Backup

Z-> b$\bar{b}$ in *pp* collisions at $\sqrt{s}$ = 8 TeV

- Z->b$\bar{b}$ is an important background for new physics searches
- Jets required to be SV-tagged, have p_T > 20 GeV, 45 < m_{jj} < 165 GeV, and 2.2 < η_{jet} < 4.2
- Uniform Gradient Boost BDT is used to discriminate Z(->b$\bar{b}$) + jet events from QCD multijet events
- Systematics impacted by theory uncertainties:
  - Recoil jet selection efficiency corrected at NLO using Z->b$\bar{b}$ events produced with aMC@NLO + Pythia for parton showers, 1.8% systematic
  - Fit repeated with subdominant backgrounds t$\bar{t}$ and W->qq' fixed to SM predictions, 1.9% systematic



measured:

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 332 \pm 46 \pm 59 \text{ pb}$$

theory:

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 272^{+9}_{-12}(\text{scale}) \pm 5(\text{PDFs}) \text{ pb}$$

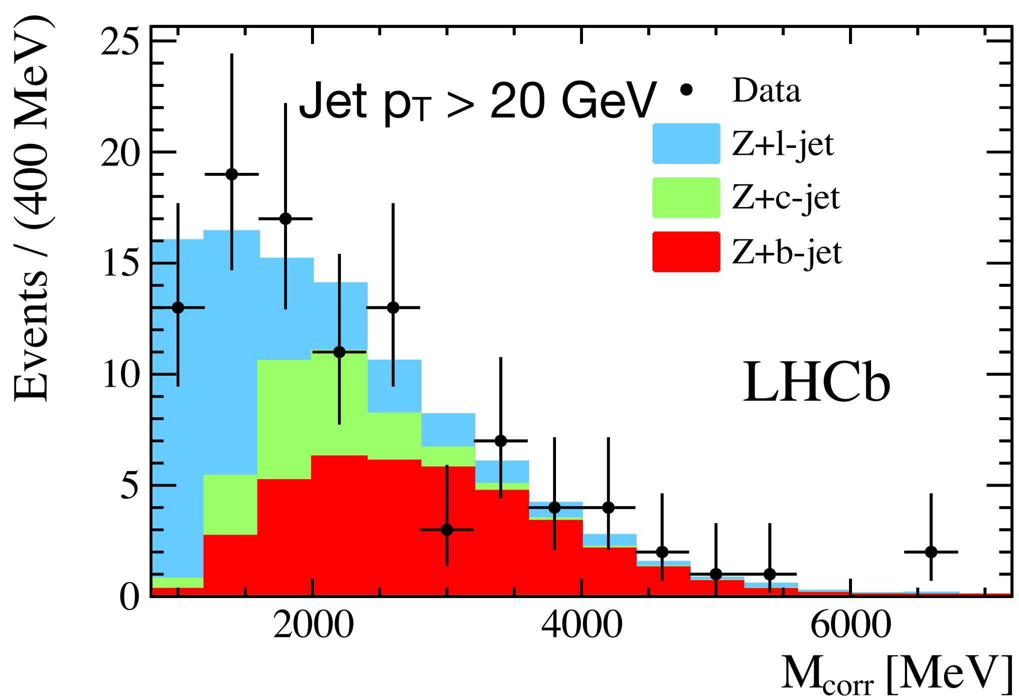
 PLB 776 (2018) 430

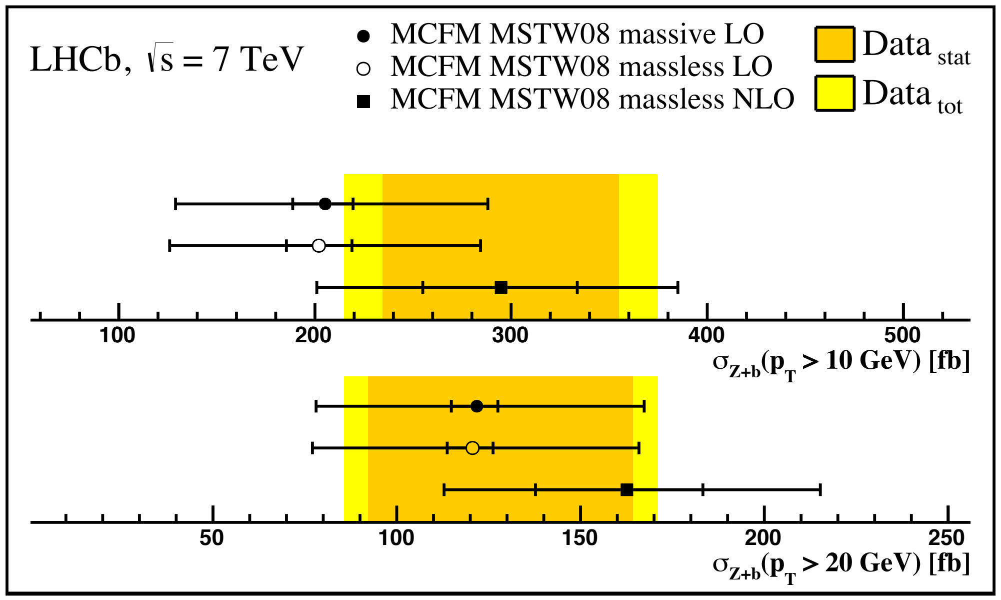
Z+b-jet cross section at √s = 7 TeV

- Sensitive to g->b$\bar{b}$ splitting, intrinsic b quark content in the proton, and an important background to constrain for Higgs and BSM studies
- Measured at LHCb with 1 fb^{-1} of data, reconstructed Z boson with a jet tagged by the TOPO SV algorithm
  - $2 < \eta(\mu) < 4.5$, $p_T(\mu) > 20$ GeV, $60 < M(\mu^+\mu^-) < 120$ GeV, $2 < \eta(\text{jet}) < 4.5$
- Measurement performed in two bins of jet pT:

Jet $p_T > 10$ GeV: $\sigma(Z/\gamma^*(\mu^+\mu^-)\text{+b-jet}) = 295 \pm 60\ (\text{stat}) \pm 51\ (\text{syst}) \pm 10\ (\text{lumi})\ \text{fb}$

Jet $p_T > 20$ GeV: $\sigma(Z/\gamma^*(\mu^+\mu^-)\text{+b-jet}) = 128 \pm 36\ (\text{stat}) \pm 22\ (\text{syst}) \pm 5\ (\text{lumi})\ \text{fb}$





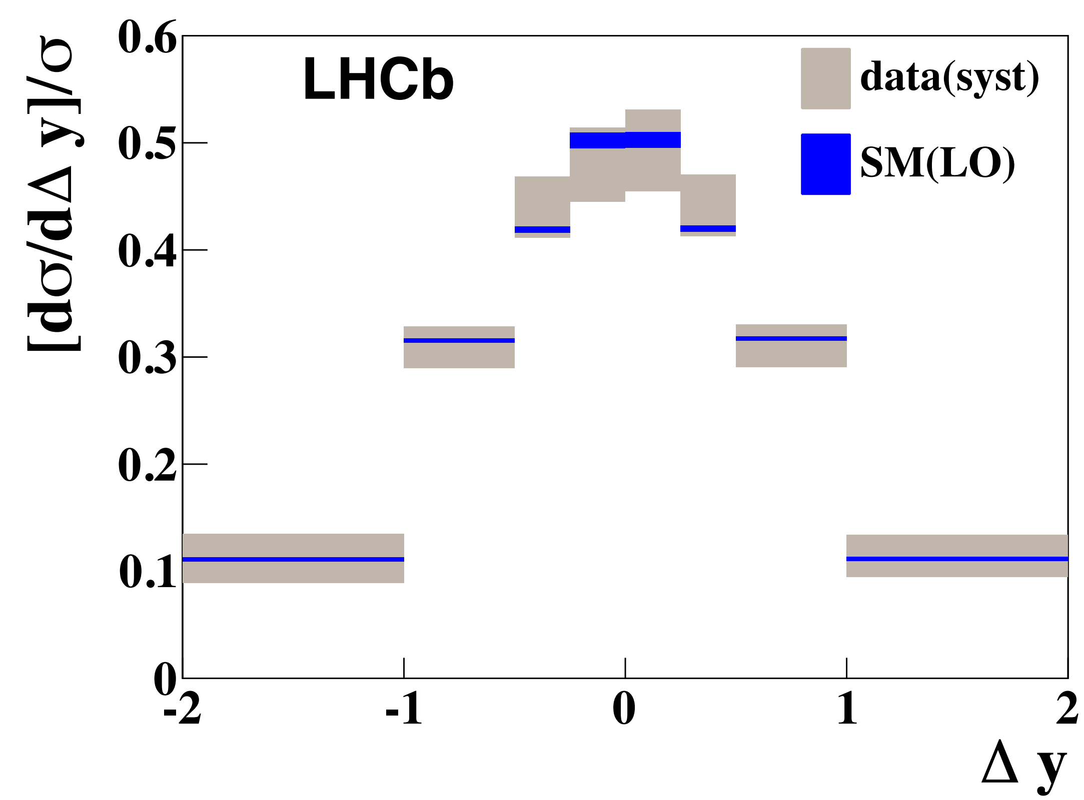
JHEP 01 (2015) 064

$b\bar{b}$ charge asymmetry in *pp* collisions at $\sqrt{s}$ = 7 TeV

$$A_C^{b\bar{b}} \equiv \frac{N(\Delta y > 0) - N(\Delta y < 0)}{N(\Delta y > 0) + N(\Delta y < 0)}$$

$$|\Delta y| = |y_b| - |y_{\bar{b}}|$$

- Precision measurements of charge asymmetries can probe BSM physics
- b-jets are tagged with the TOPO algorithm, $2 < \eta_{jet} < 4$, $E_T > 20$ GeV, $\Delta\phi > 2.6$
- Charge tagging performed by requiring one of the tracks in the SV to be a muon
- Measurements consistent with SM predictions:



$$A_C^{b\bar{b}}(40 < M_{b\bar{b}} < 75\,\mathrm{GeV}/c^2) = 0.4 \pm 0.4\,(\mathrm{stat}) \pm 0.3\,(\mathrm{syst})\%$$
$$A_C^{b\bar{b}}(75 < M_{b\bar{b}} < 105\,\mathrm{GeV}/c^2) = 2.0 \pm 0.9\,(\mathrm{stat}) \pm 0.6\,(\mathrm{syst})\%$$
$$A_C^{b\bar{b}}(M_{b\bar{b}} > 105\,\mathrm{GeV}/c^2) = 1.6 \pm 1.7\,(\mathrm{stat}) \pm 0.6\,(\mathrm{syst})\%$$

PRL 113, 082003 (2014)

TOPO algorithm as a b-jet tagger

- BDT used in the LHCb trigger to identify secondary vertices consistent with b-hadron decays
- Builds 2, 3, and 4-track SVs

TOPO BDT inputs:

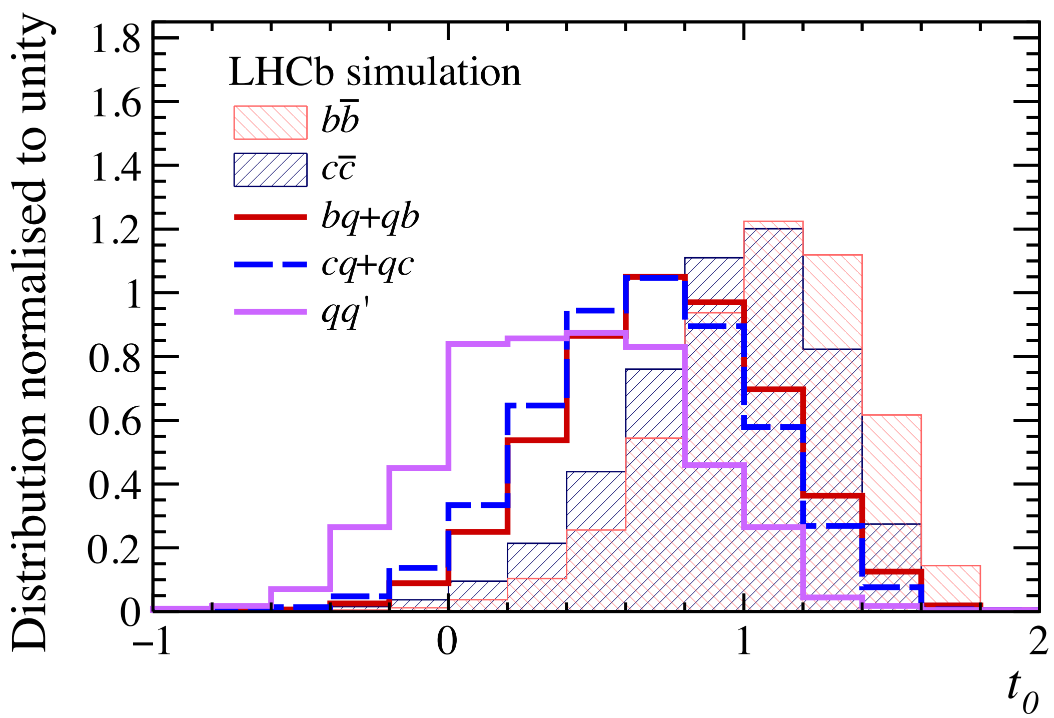
- SV mass
- SV corrected mass
- Sum of the p_T of the SV tracks
- Maximum distance of closest approach between the SV tracks
- Sum of the difference in χ^2 when performing a PV fit with and without each track in the SV
- SV flight distance χ^2
- Minimum p_T of the SV tracks

Also input to BDTs for the SV-tagger algorithm

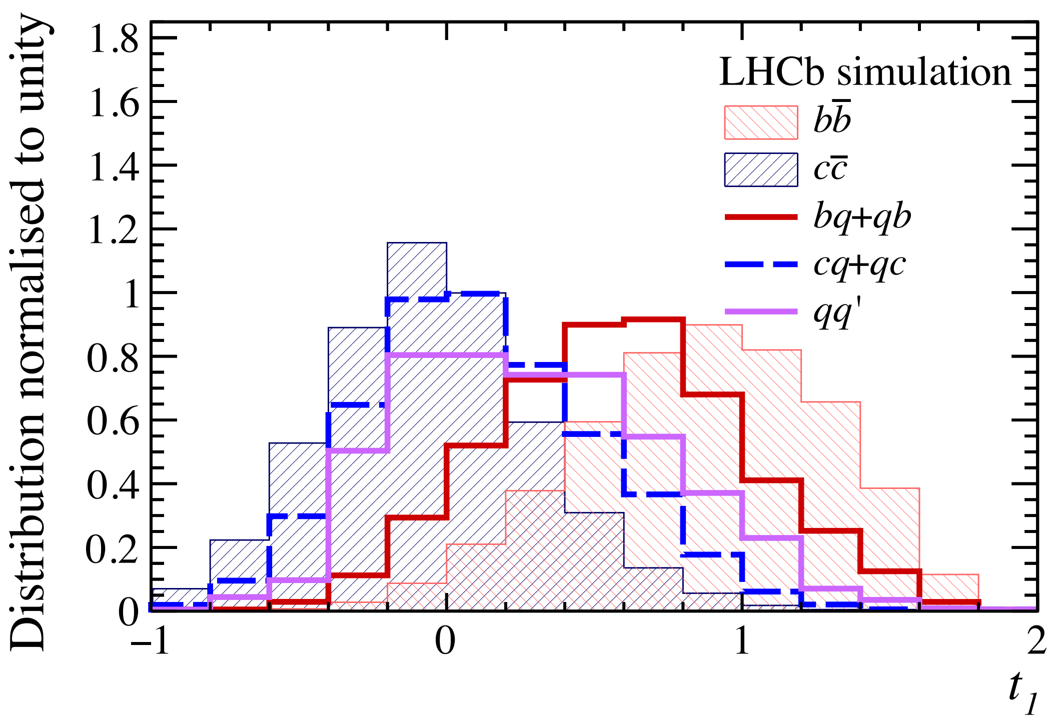
JINST 10, P06013 (2015)

t_0 and t_1 templates from simulation

t_0 = $BDT_{bc|udsg}$(jet 0) + $BDT_{bc|udsg}$(jet 1)



t_1 = $BDT_{b|c}$(jet 0) + $BDT_{b|c}$(jet 1)



- $b\bar{b}$, $c\bar{c}$, qq’ templates constructed from simulated events
- bq, qb, cq,qc templates constructed from convolutions of single-jet 2D BDT templates (BDT(bc|udsg), BDT(b|c))
- Templates constructed for the following bins of [jet 0, jet 1] pT: [20,30] GeV, [30,40] GeV, [40,50] GeV, [50,60] GeV, and >60 GeV