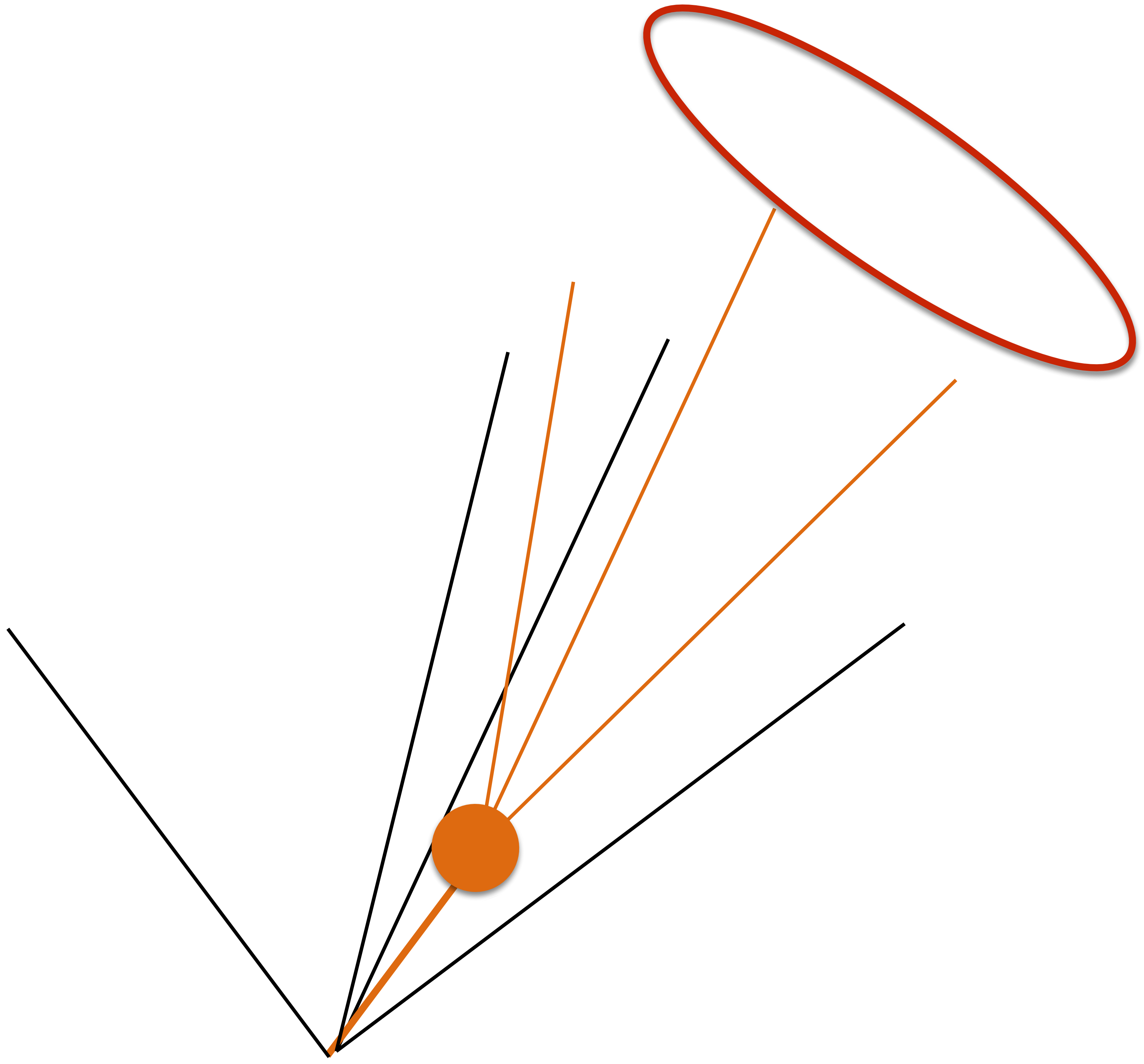
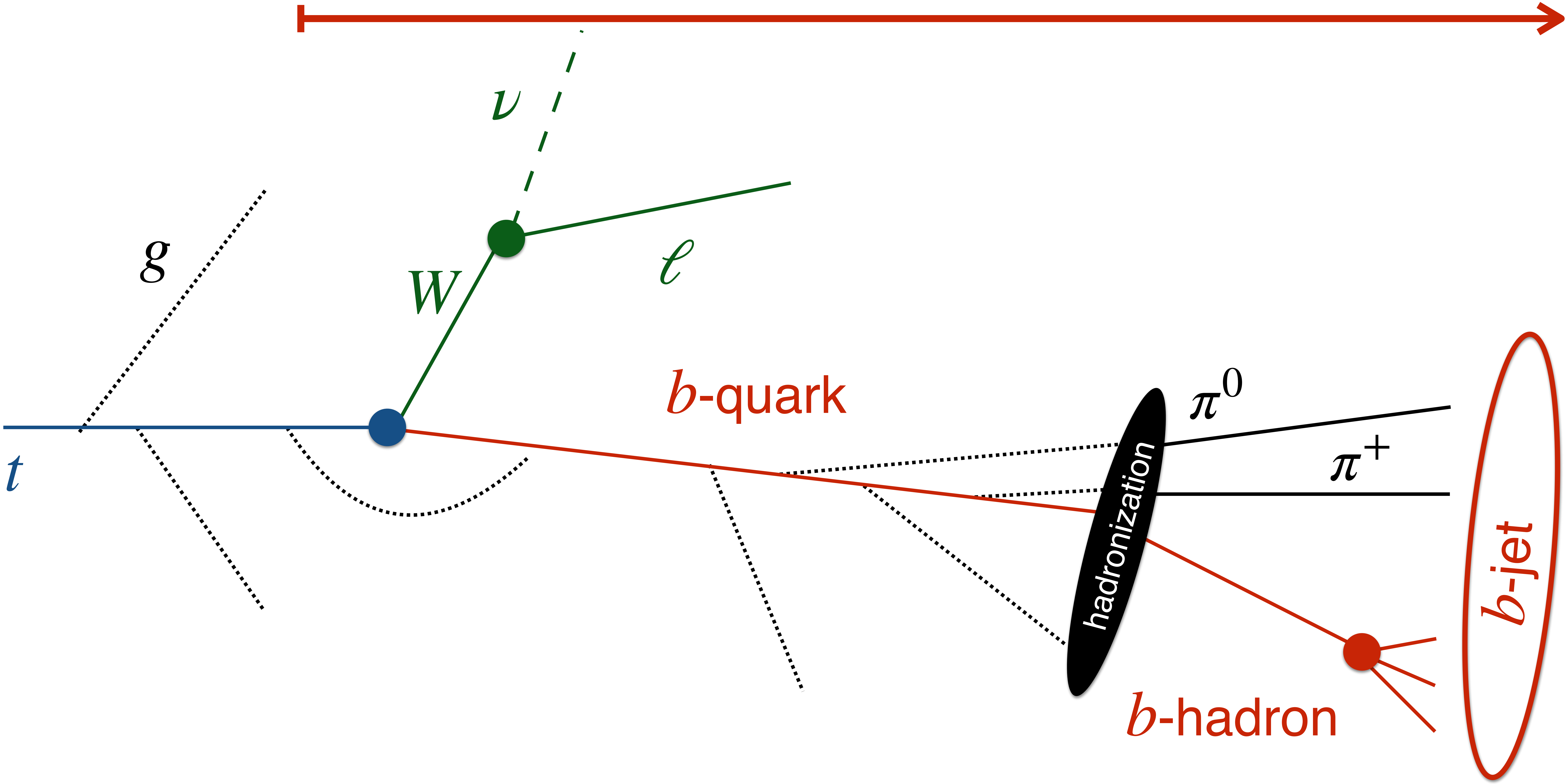
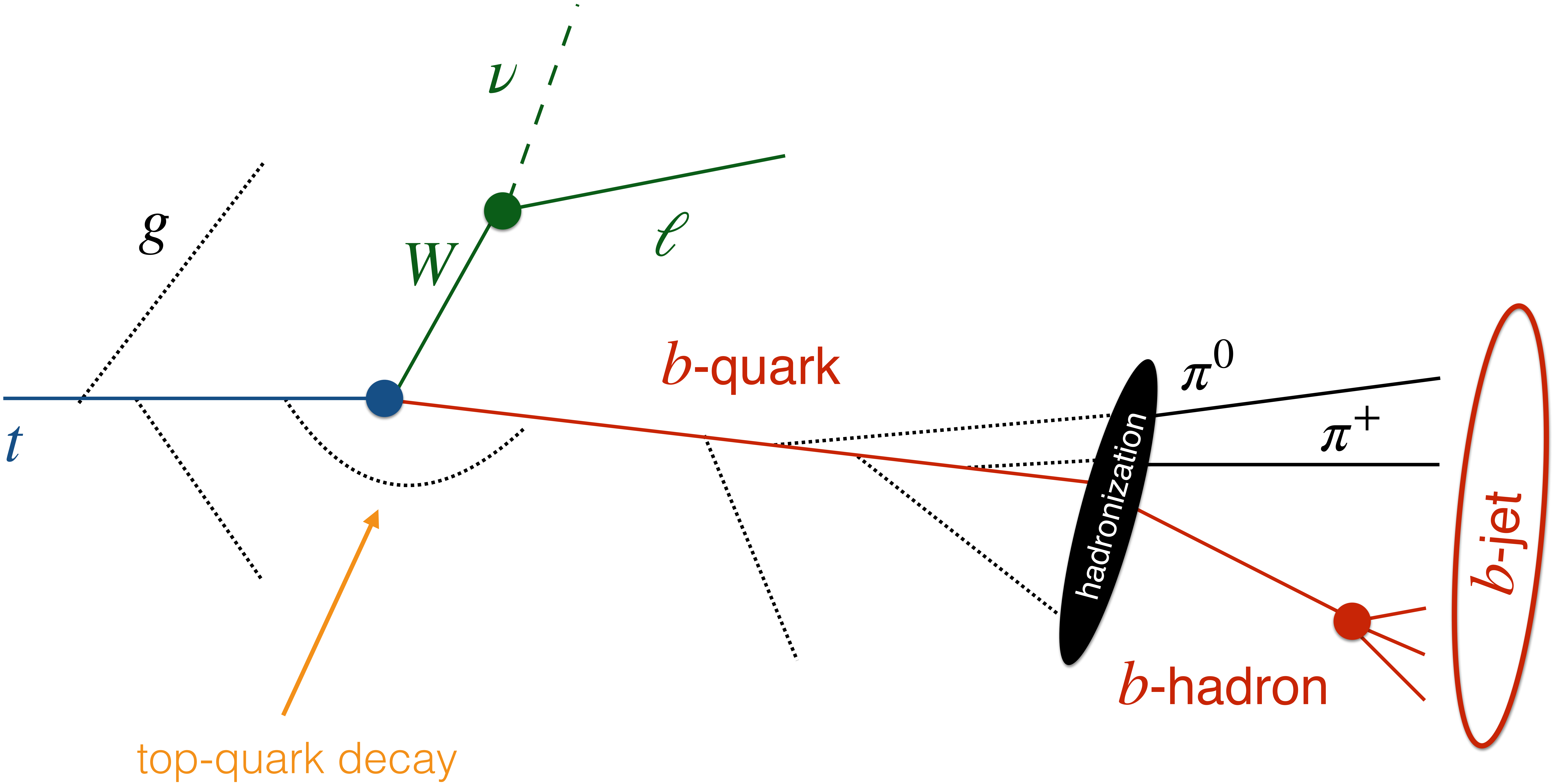
recent (and future) measurements of heavy-quark fragmentation

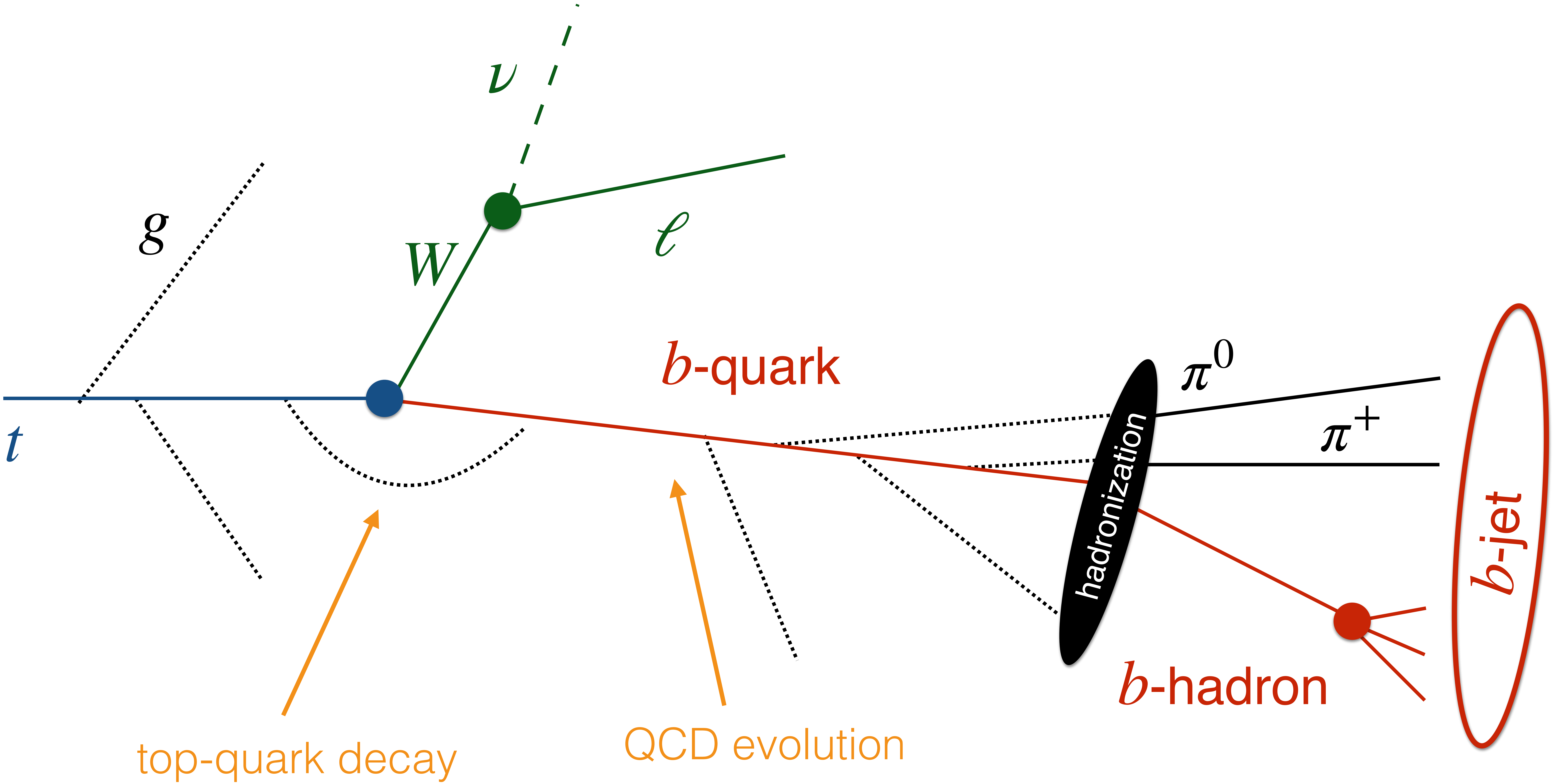
Chris Pollard
University of Warwick



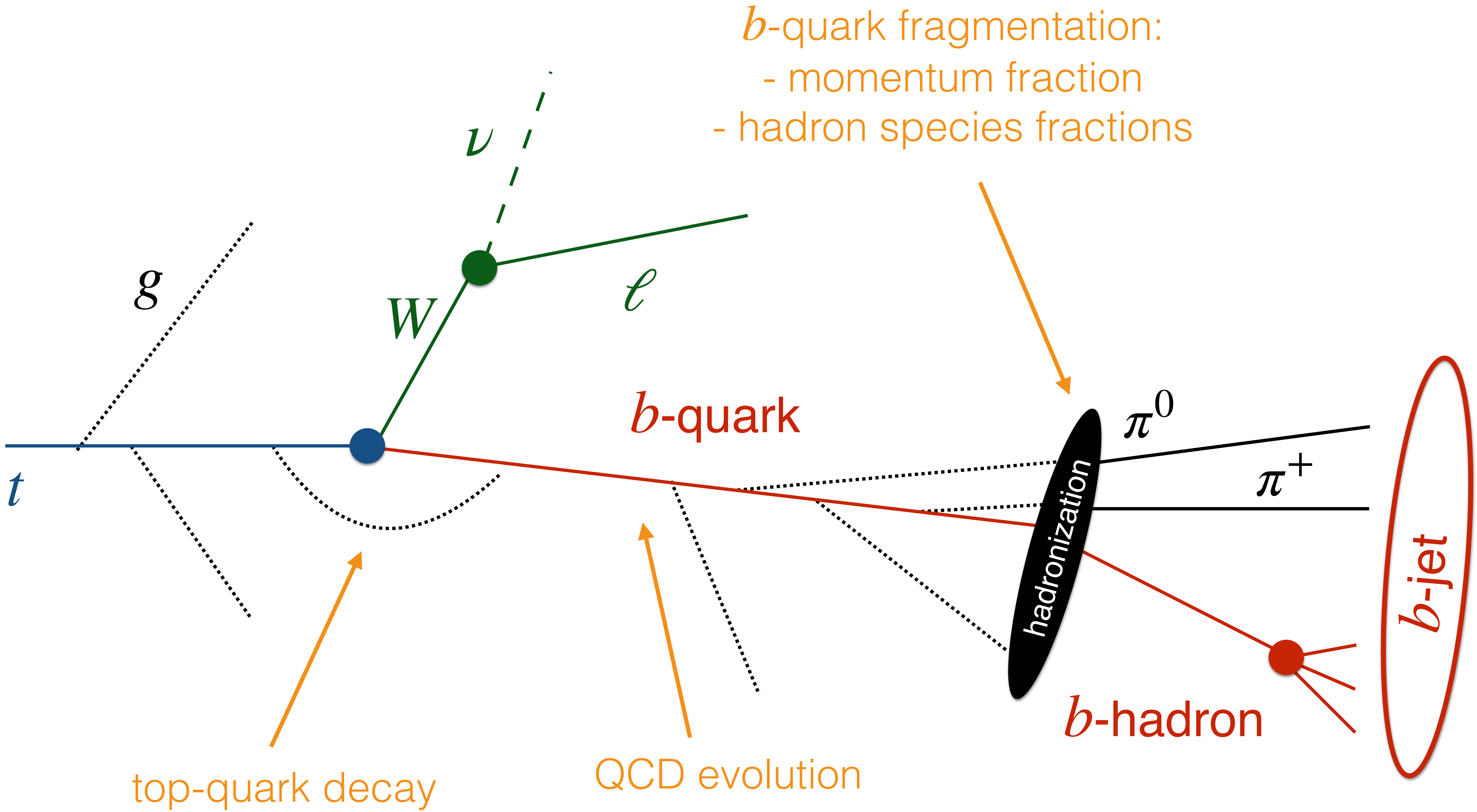
we need to understand this whole process!

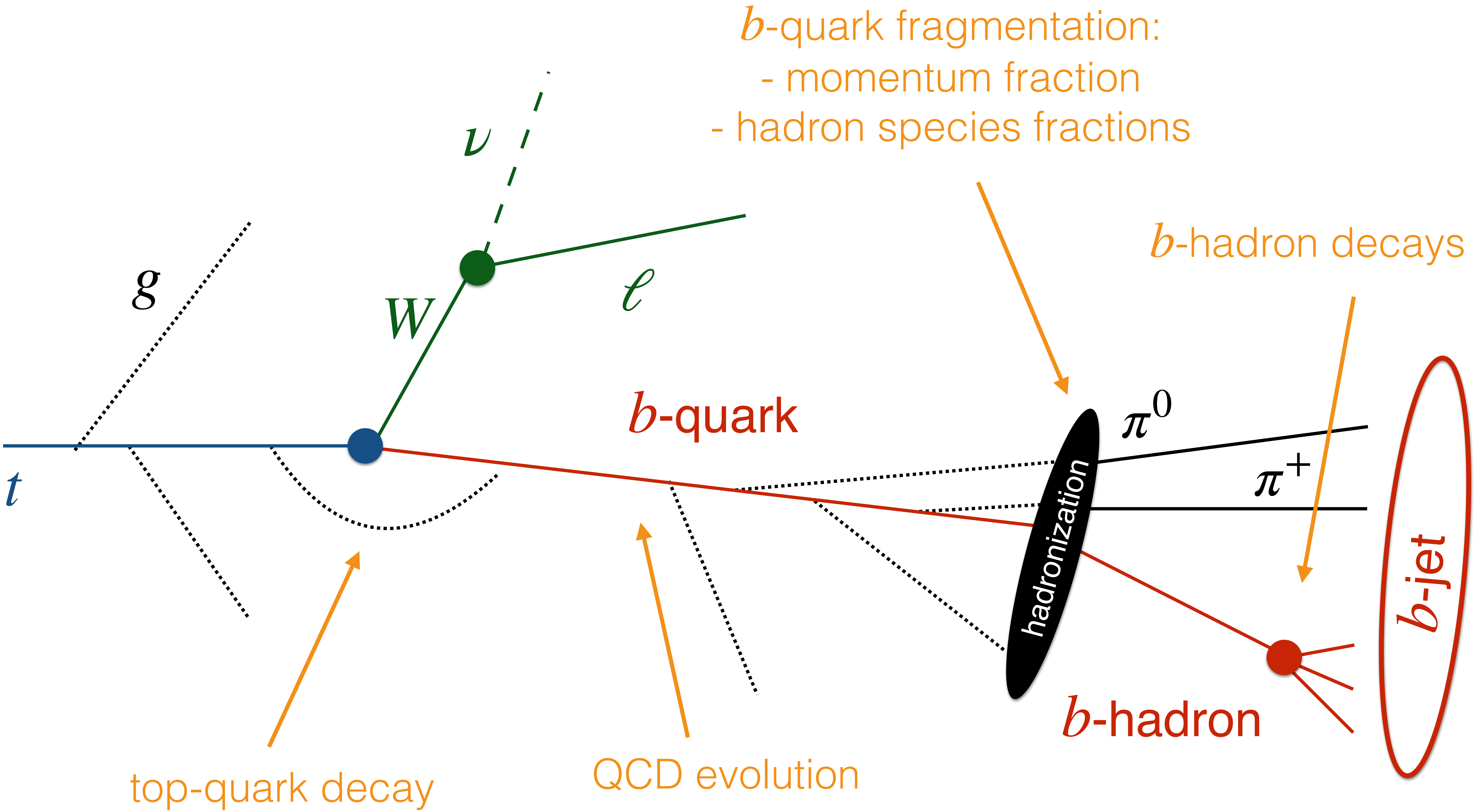



ν
g
W
ℓ
b-quark
π^0
π^+
t
hadronization
b-jet
b-hadron
top-quark decay



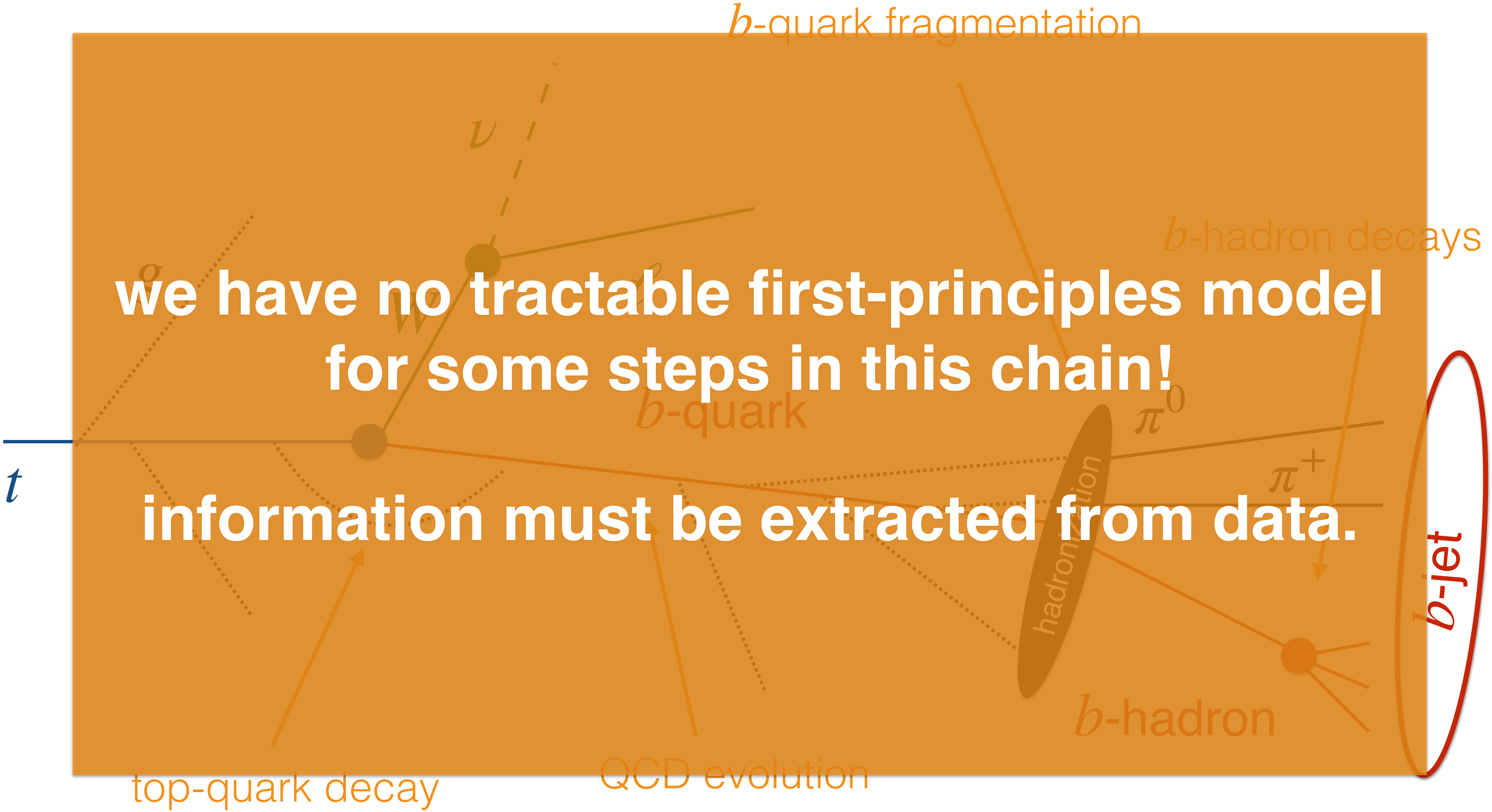
ν
g
W
ℓ
t
b-quark
π^0
π^+
hadronization
b-jet
b-hadron
top-quark decay
QCD evolution


b-quark fragmentation:
- momentum fraction
- hadron species fractions
ν
g
W
ℓ
b-quark
π0
π+
t
hadronization
b-jet
b-hadron
top-quark decay
QCD evolution

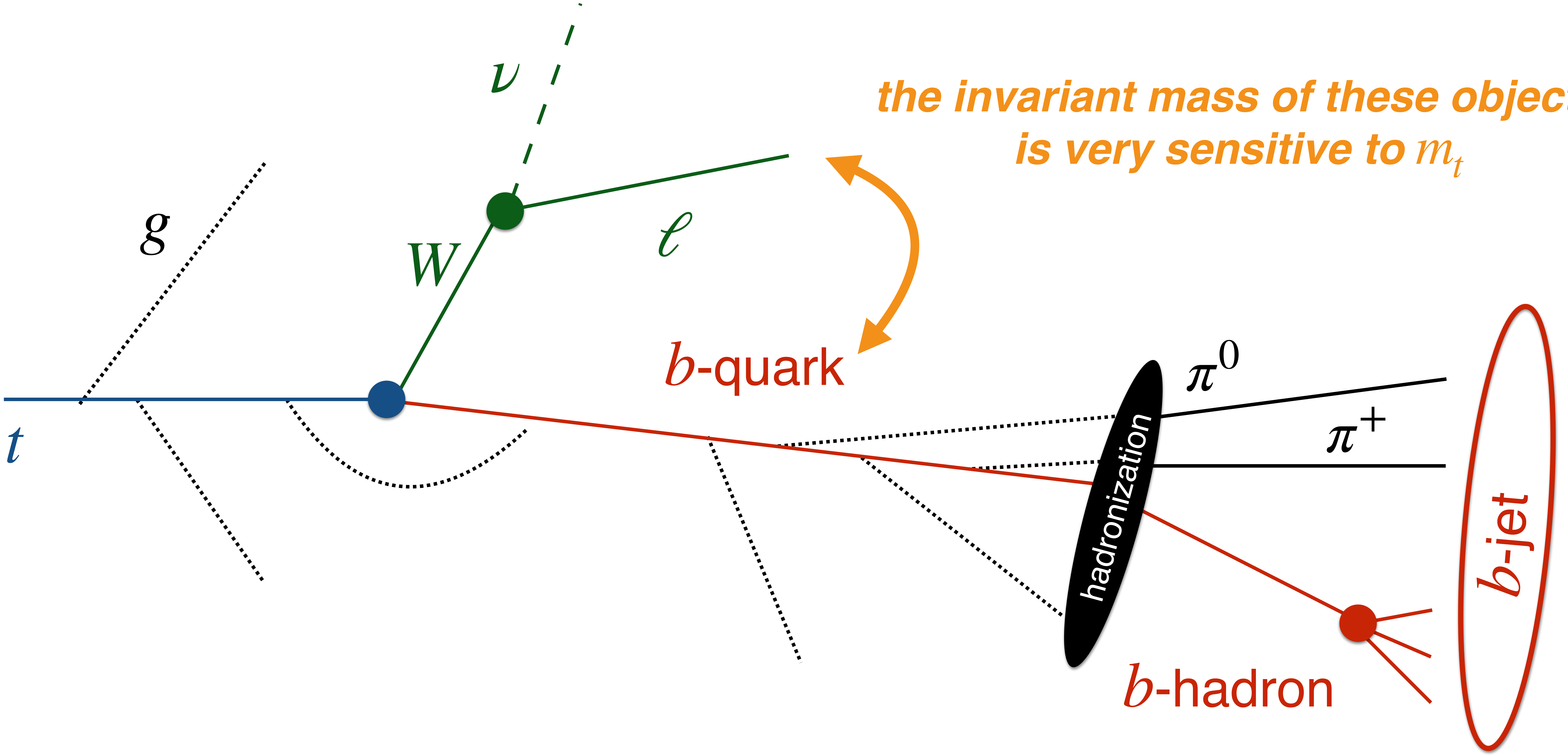

b-quark fragmentation:
- momentum fraction
- hadron species fractions
b-hadron decays
g
ν
W
ℓ
t
b-quark
π0
π+
hadronization
b-jet
b-hadron
top-quark decay
QCD evolution

we have no tractable first-principles model for some steps in this chain!

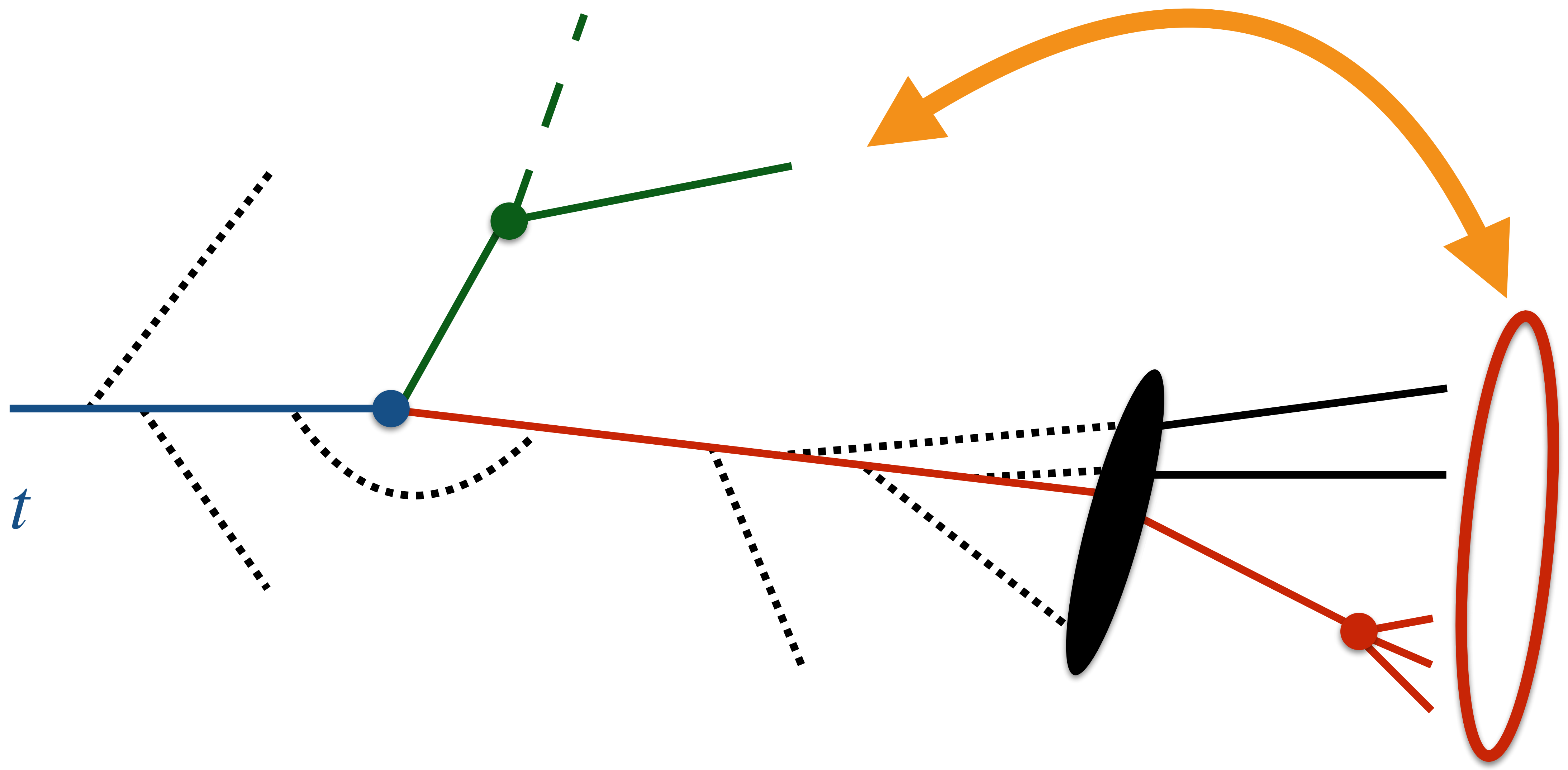
information must be extracted from data.

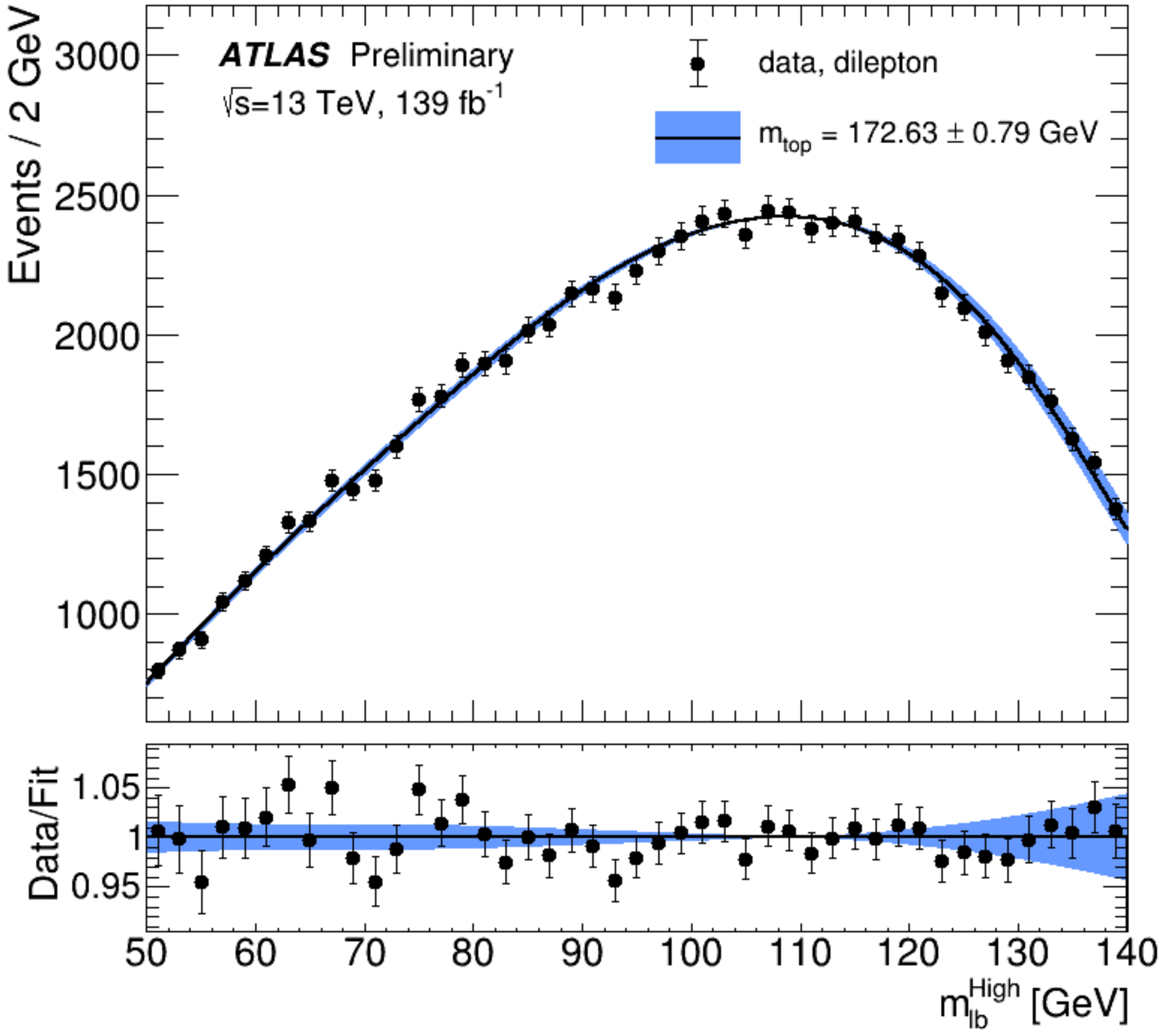


why?



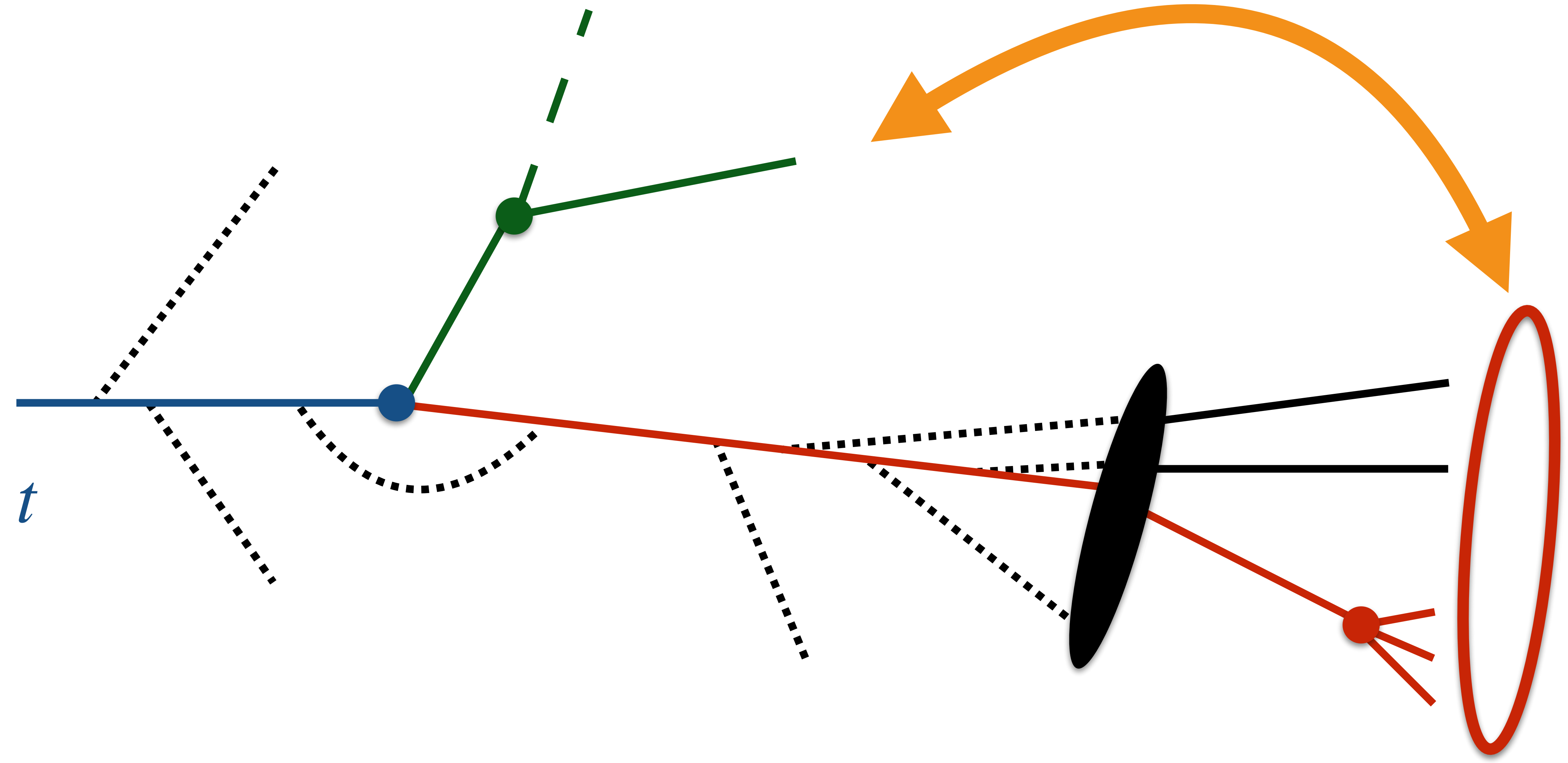
the invariant mass of these objects is very sensitive to m_t

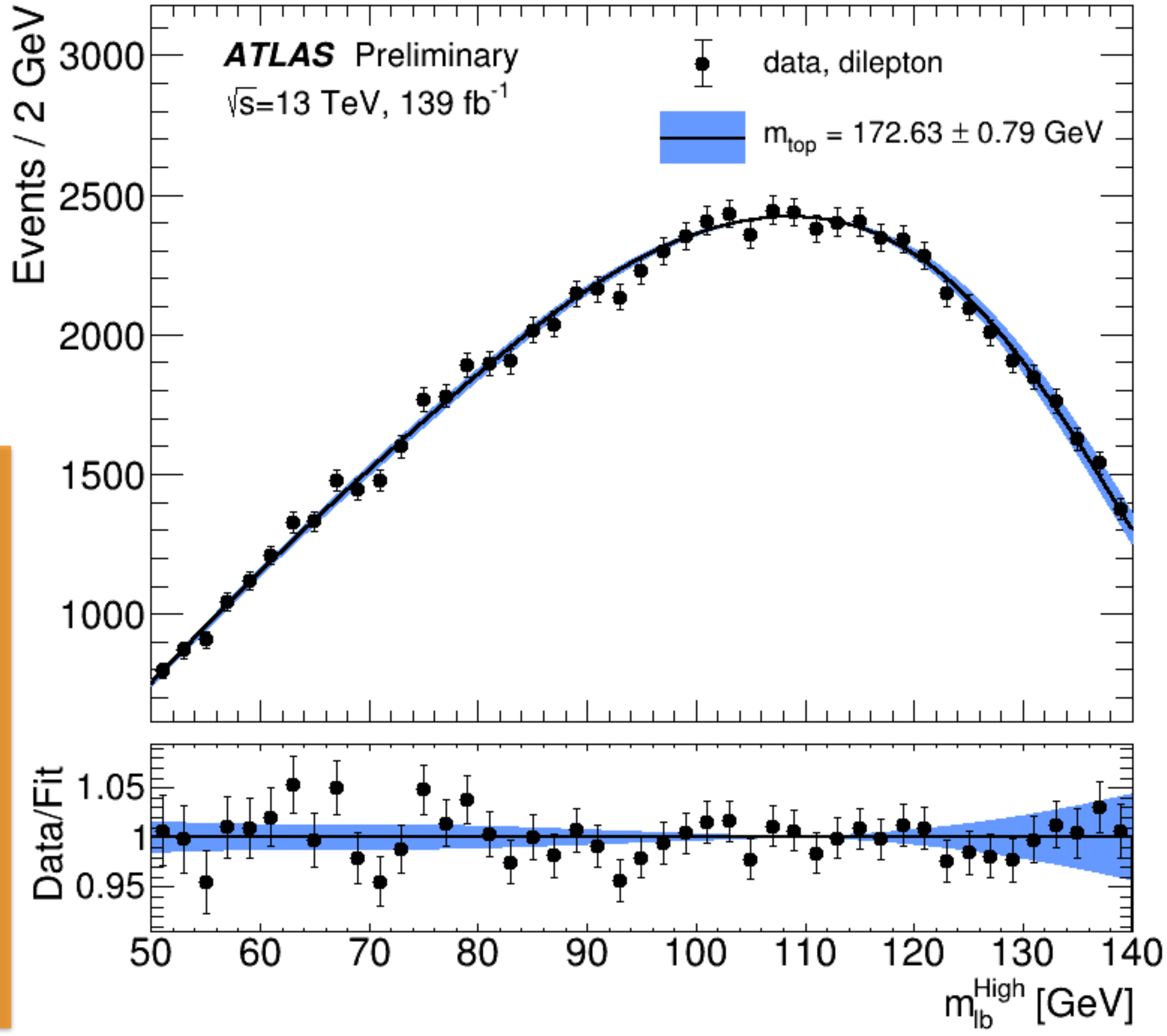




ATLAS recently measured m_t via the invariant mass of the W lepton and the b-jet

$$m_t = 172.63 \pm 0.79 \text{ GeV}$$





ATLAS recently measured m_t via the invariant mass of the W lepton and the b-jet

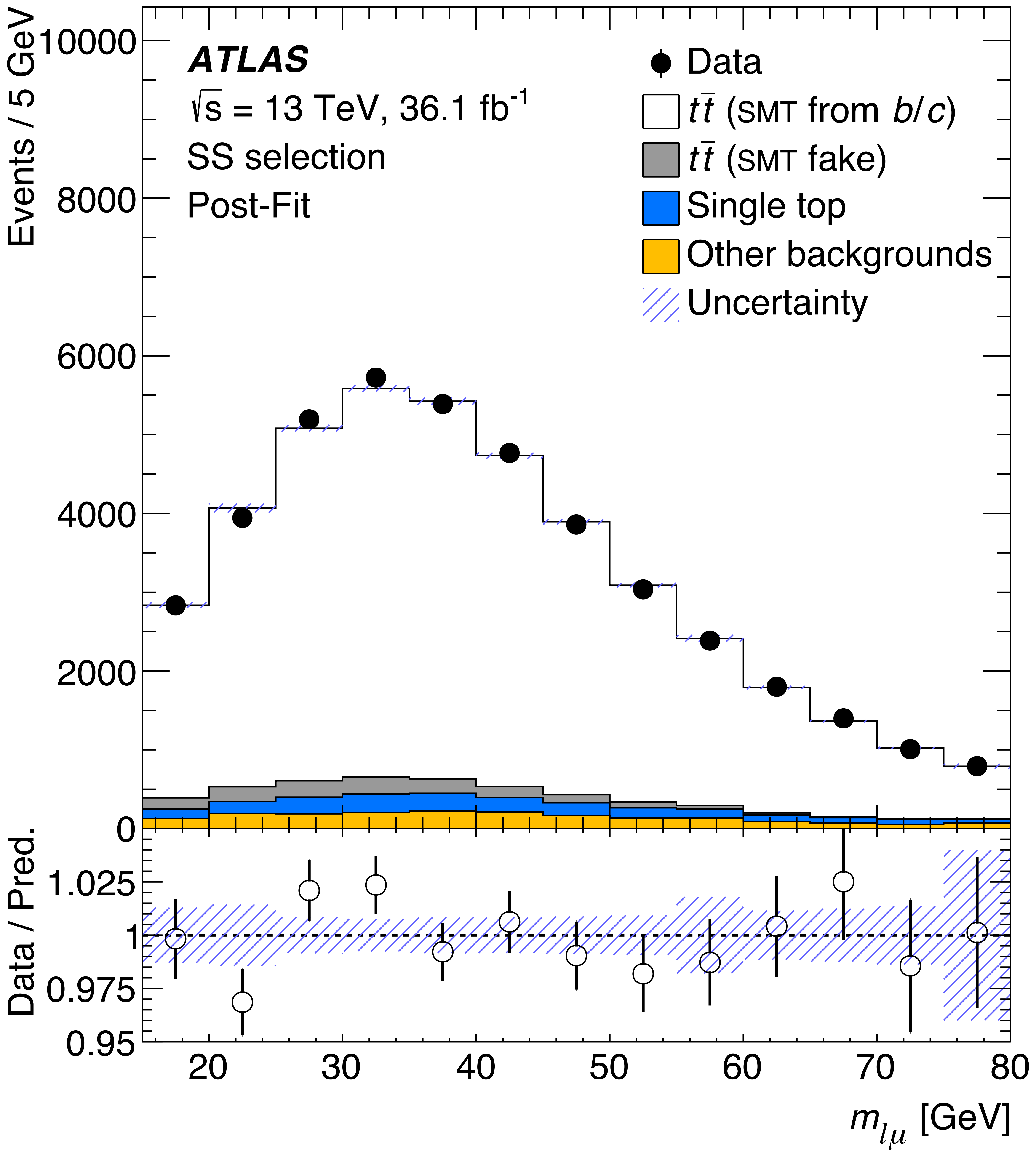
$$m_t = 172.63 \pm 0.79 \text{ GeV}$$

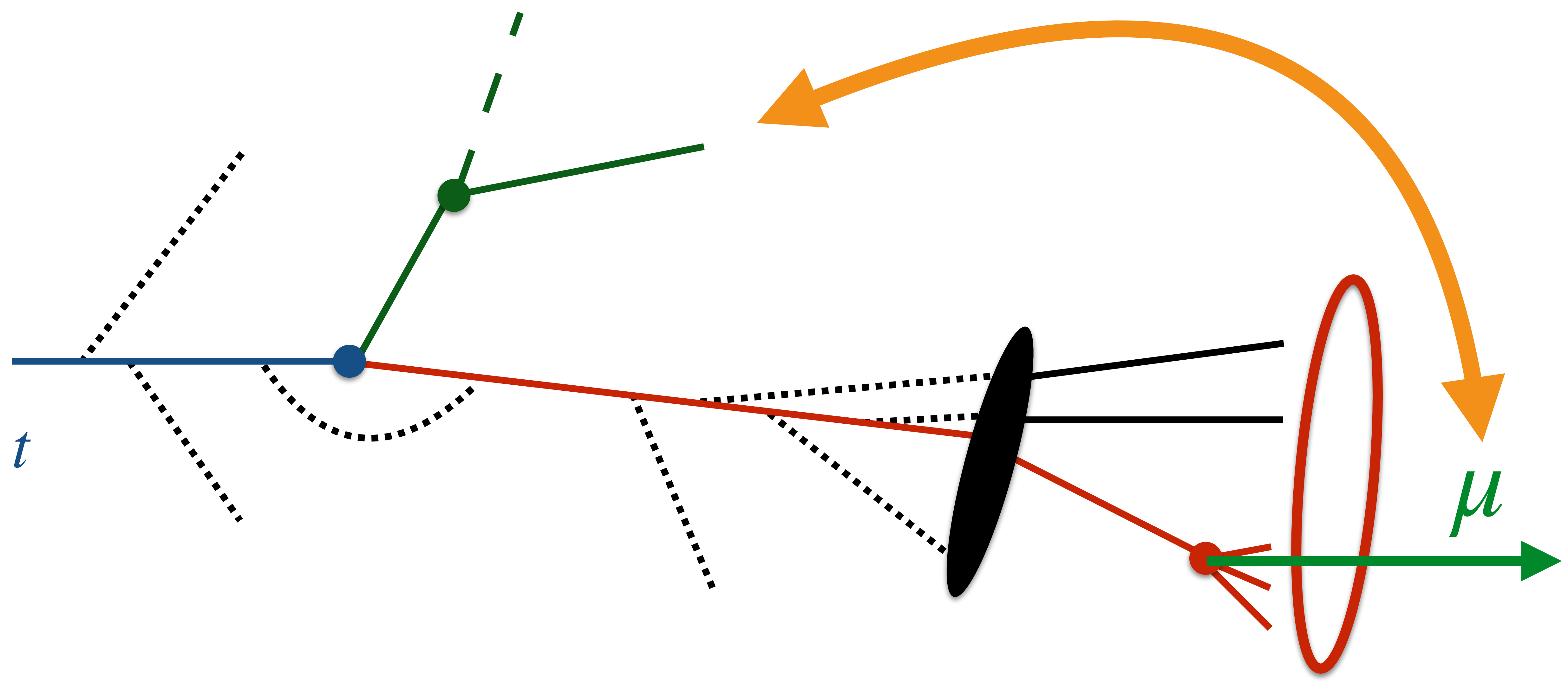
sensitive to

high-order effects in top-quark decay,

“out of cone” radiation off of the b-quark,

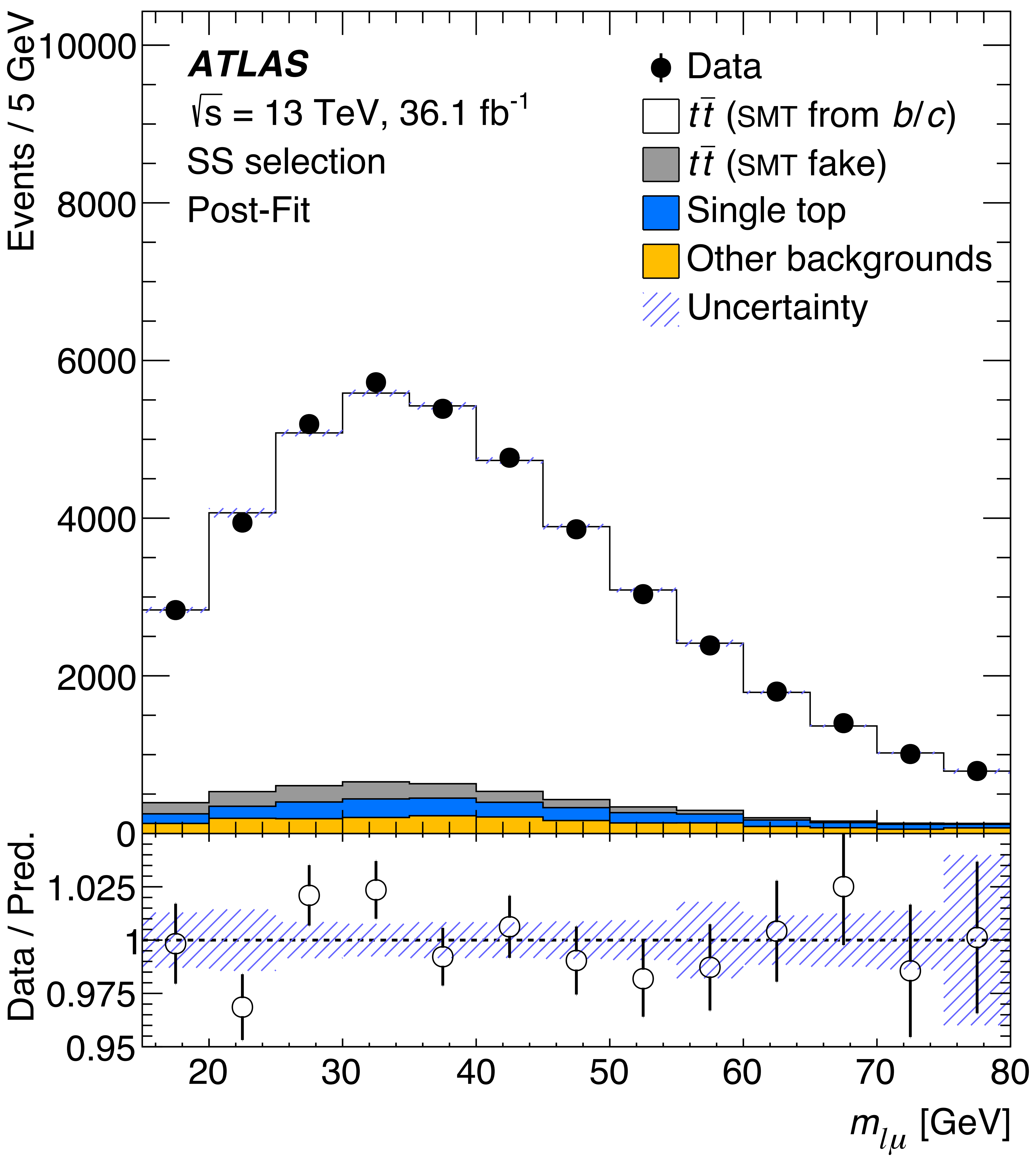
“in cone” radiation via the jet response.

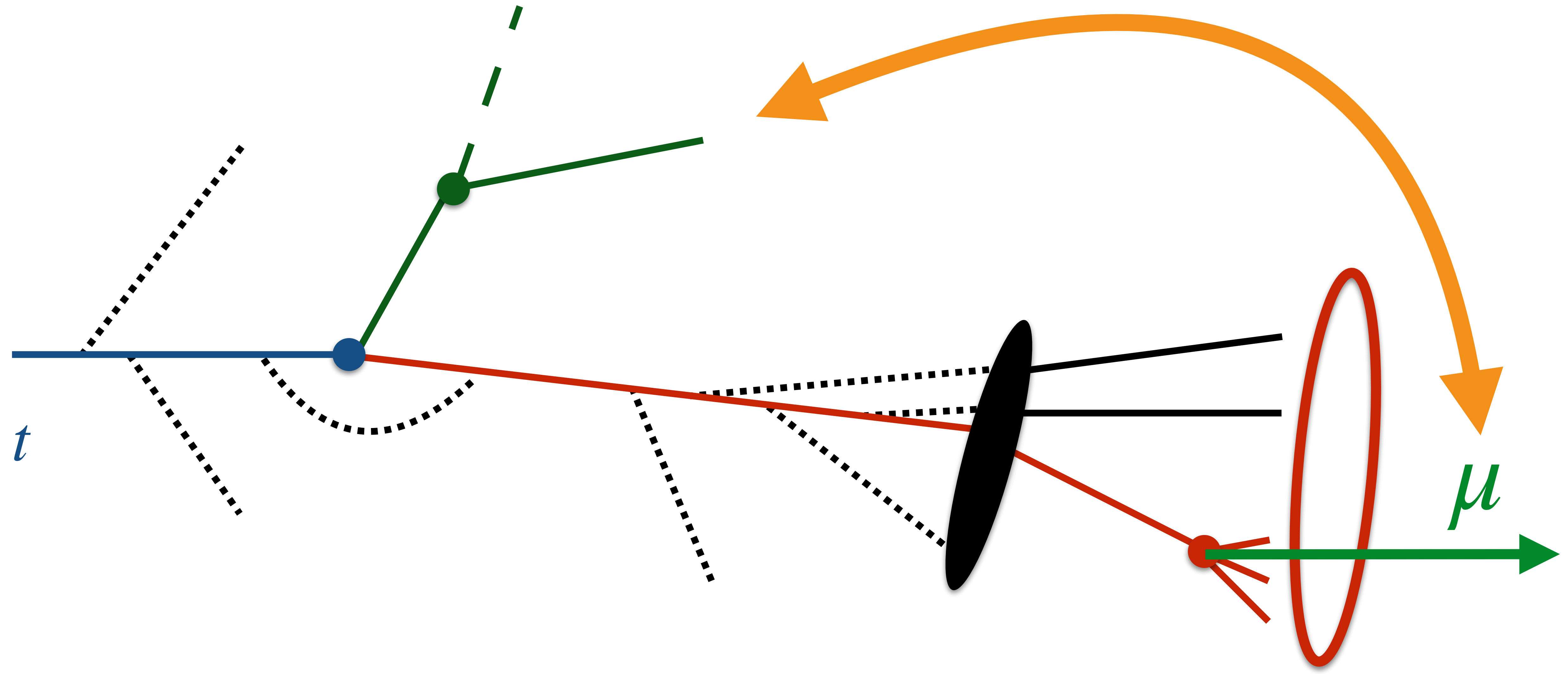




ATLAS also recently measured m_t *via the invariant mass of the* W *lepton and a lepton from the* b*-hadron decay*

$$m_t = 174.71 \pm 0.81\ \text{GeV}$$





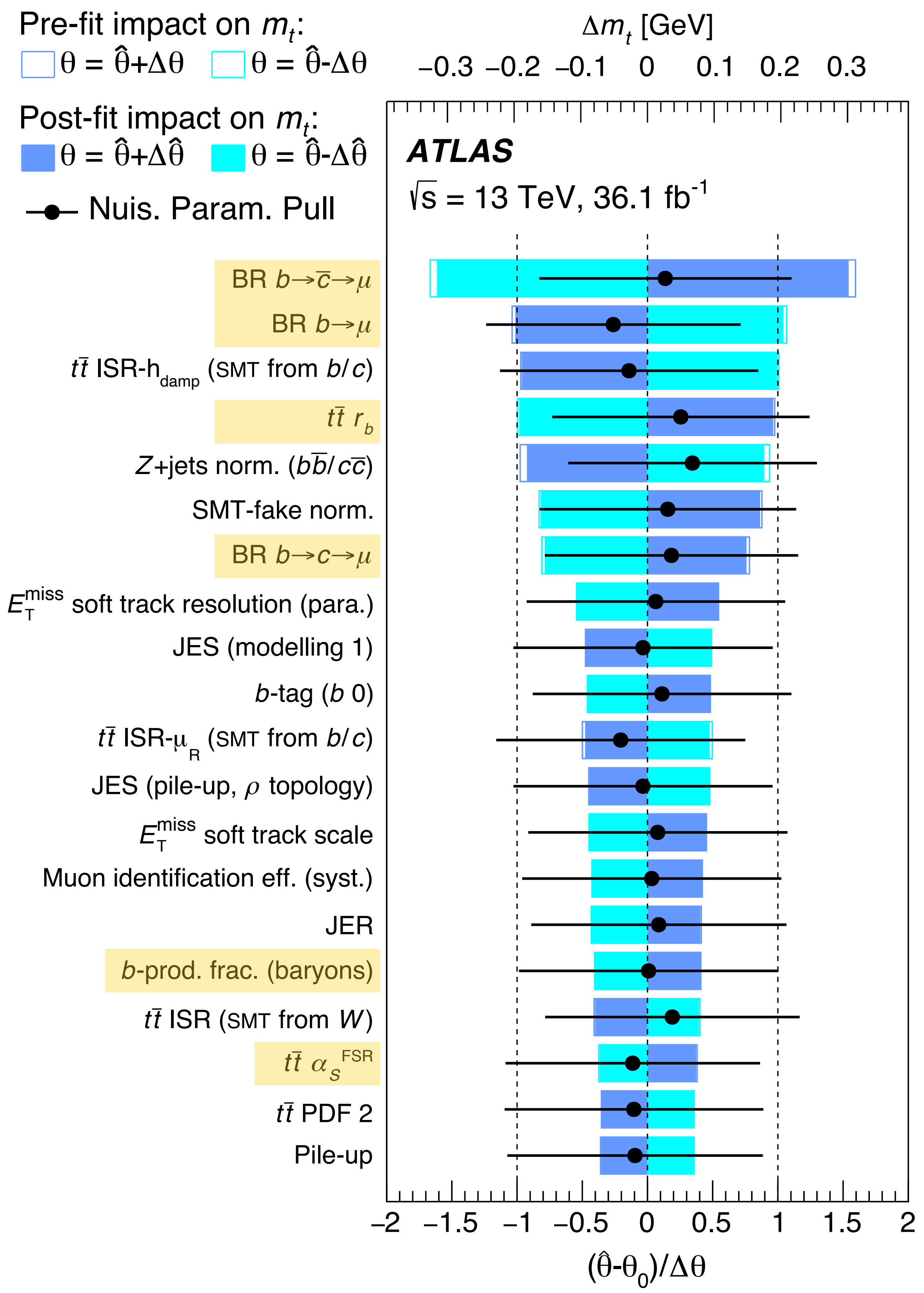
sensitive to

high-order effects in top-quark decay,

“out of cone” radiation off of the b-quark,

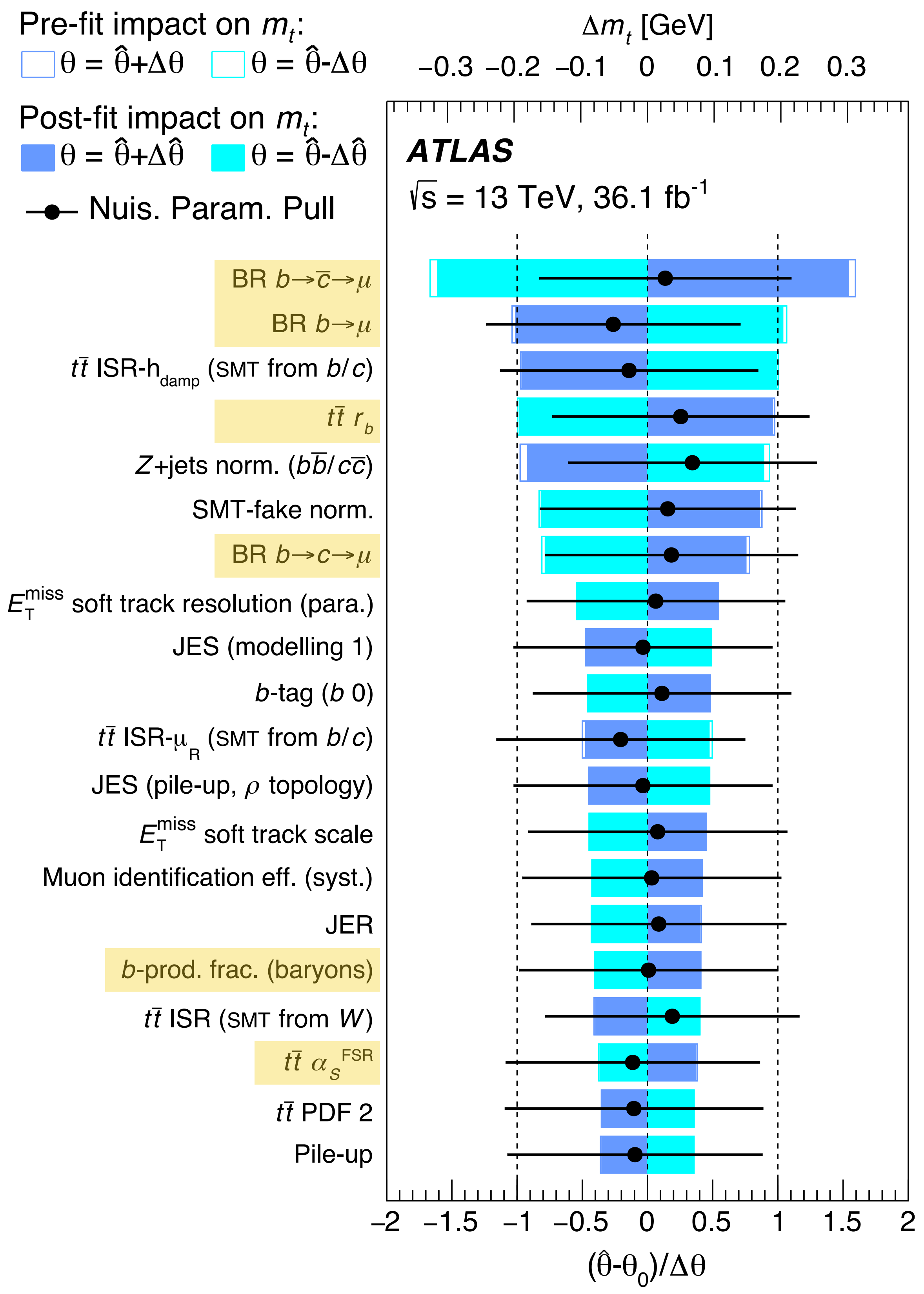
b-hadron production and decay fractions.

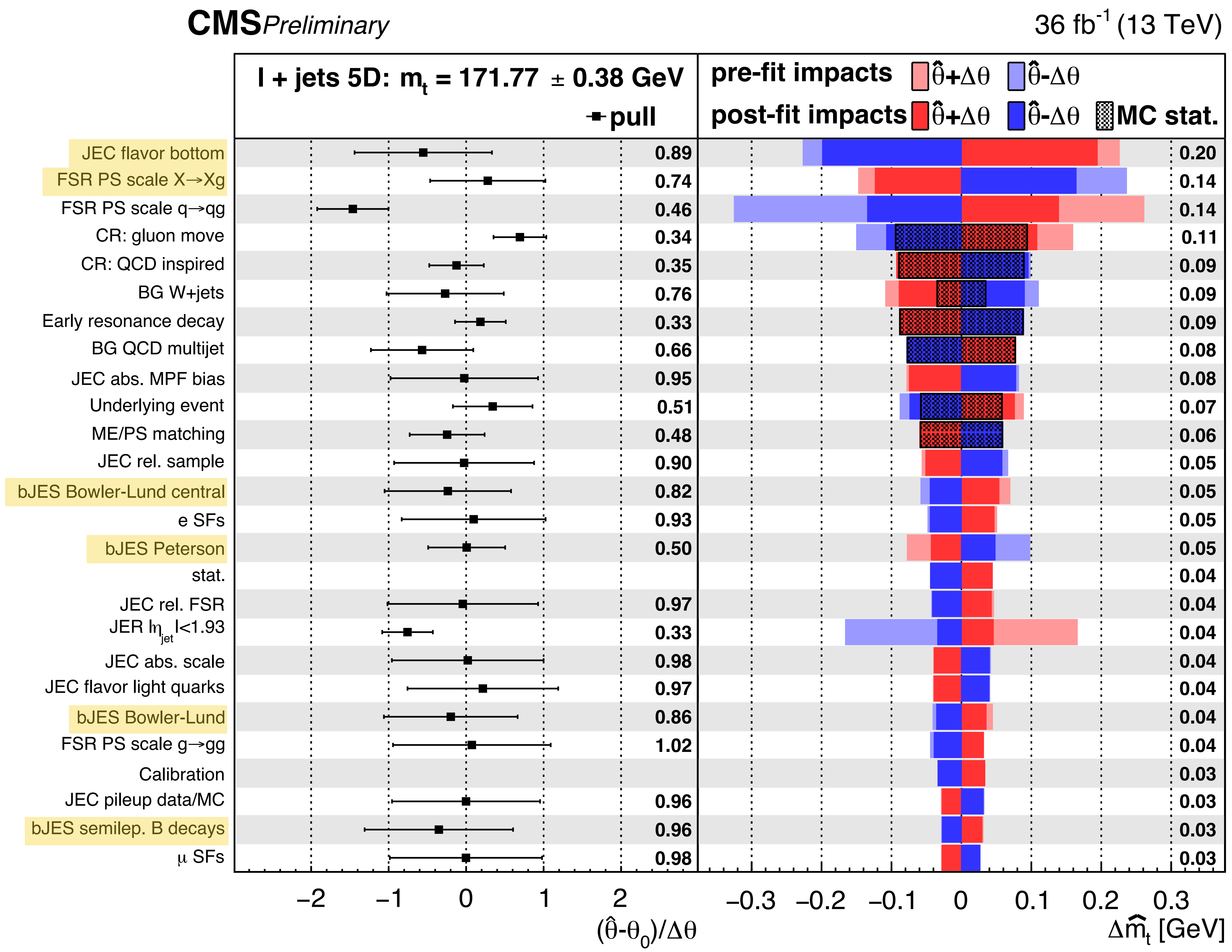
	m_{top} [GeV]
Result	172.63
Statistics	0.20
Method	0.05 ± 0.04
Matrix-element matching	0.35 ± 0.07
Parton shower and hadronisation	0.08 ± 0.05
Initial- and final-state QCD radiation	0.20 ± 0.02
Underlying event	0.06 ± 0.10
Colour reconnection	0.29 ± 0.07
Parton distribution function	0.02 ± 0.00
Single top modelling	0.03 ± 0.01
Background normalisation	0.01 ± 0.02
Jet energy scale	0.38 ± 0.02
b-jet energy scale	0.14 ± 0.02
Jet energy resolution	0.05 ± 0.02
Jet vertex tagging	0.01 ± 0.01
b-tagging	0.04 ± 0.01
Leptons	0.12 ± 0.02
Pile-up	0.06 ± 0.01
Recoil effect	0.37 ± 0.09
Total systematic uncertainty (without recoil)	0.67 ± 0.05
Total systematic uncertainty (with recoil)	0.77 ± 0.06
Total uncertainty (without recoil)	0.70 ± 0.05
Total uncertainty (with recoil)	0.79 ± 0.06



most dominant uncertainties come from QCD modeling in the top-quark decay, radiation/hadronization of the b-quark, and b-hadron decays.

	m_{top} [GeV]
Result	172.63
Statistics	0.20
Method	0.05 ± 0.04
Matrix-element matching	0.35 ± 0.07
Parton shower and hadronisation	0.08 ± 0.05
Initial- and final-state QCD radiation	0.20 ± 0.02
Underlying event	0.06 ± 0.10
Colour reconnection	0.29 ± 0.07
Parton distribution function	0.02 ± 0.00
Single top modelling	0.03 ± 0.01
Background normalisation	0.01 ± 0.02
Jet energy scale	0.38 ± 0.02
b-jet energy scale	0.14 ± 0.02
Jet energy resolution	0.05 ± 0.02
Jet vertex tagging	0.01 ± 0.01
b-tagging	0.04 ± 0.01
Leptons	0.12 ± 0.02
Pile-up	0.06 ± 0.01
Recoil effect	0.37 ± 0.09
Total systematic uncertainty (without recoil)	0.67 ± 0.05
Total systematic uncertainty (with recoil)	0.77 ± 0.06
Total uncertainty (without recoil)	0.70 ± 0.05
Total uncertainty (with recoil)	0.79 ± 0.06





most dominant uncertainties come from QCD modeling in the top-quark decay, radiation/hadronization of the b-quark, and b-hadron decays.

CMS see a very similar picture

more generally… other experimental results need precision!

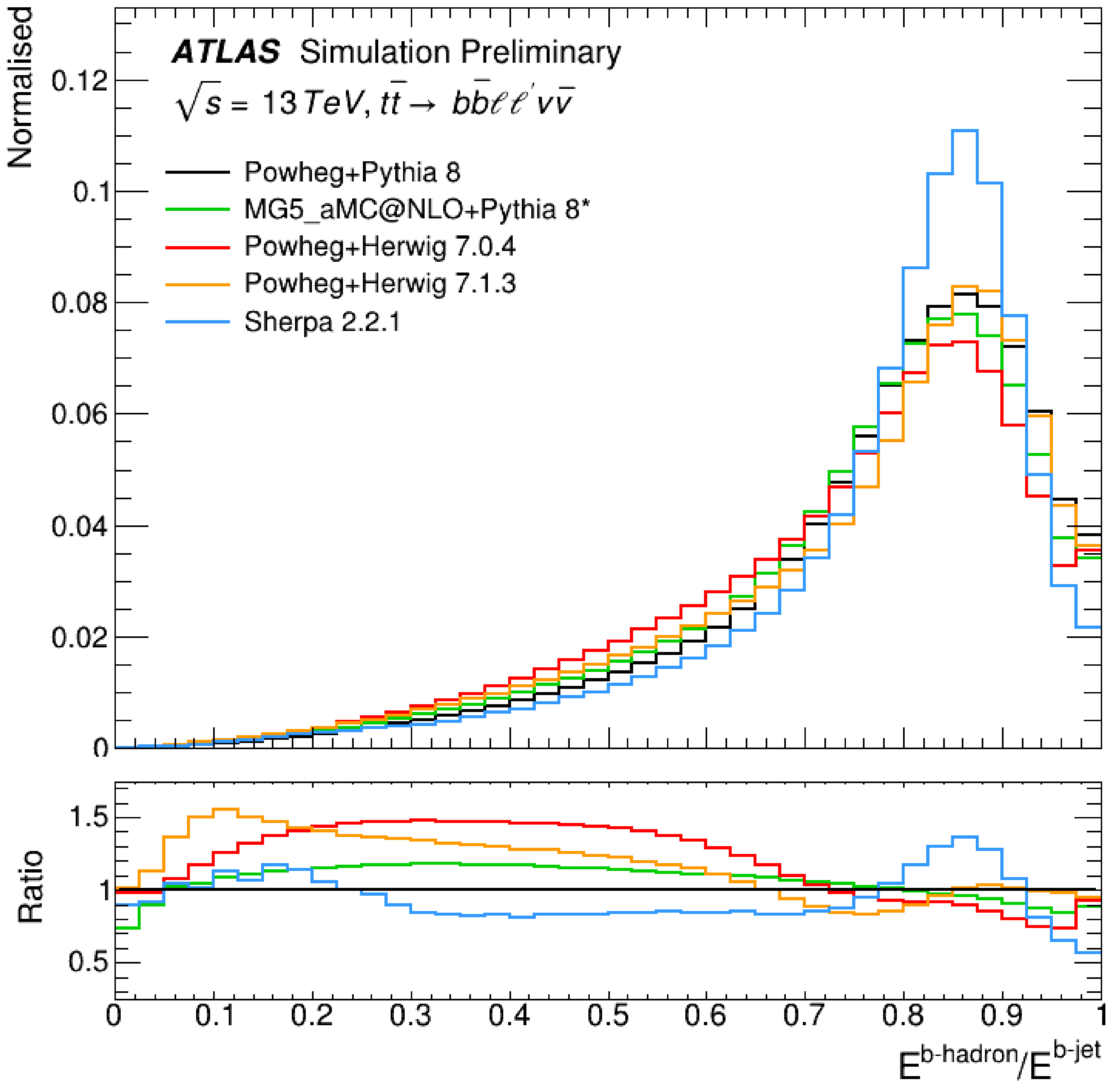
b*-tagging efficiency and* b*-jet response are very sensitive to fragmentation.*

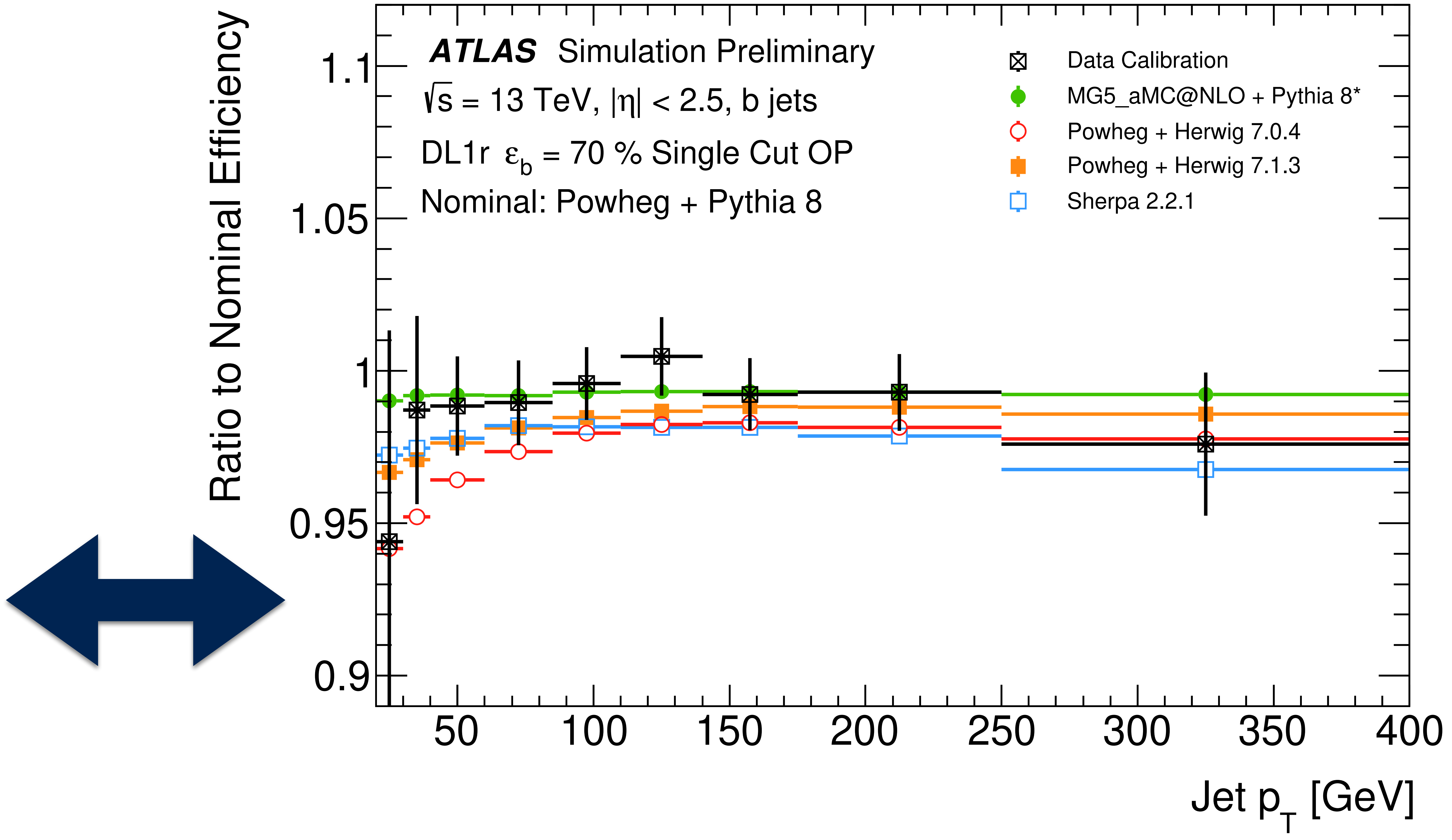
(same is true for charm)

more generally… other experimental results need precision!

b-tagging efficiency and b-jet response are very sensitive to fragmentation.

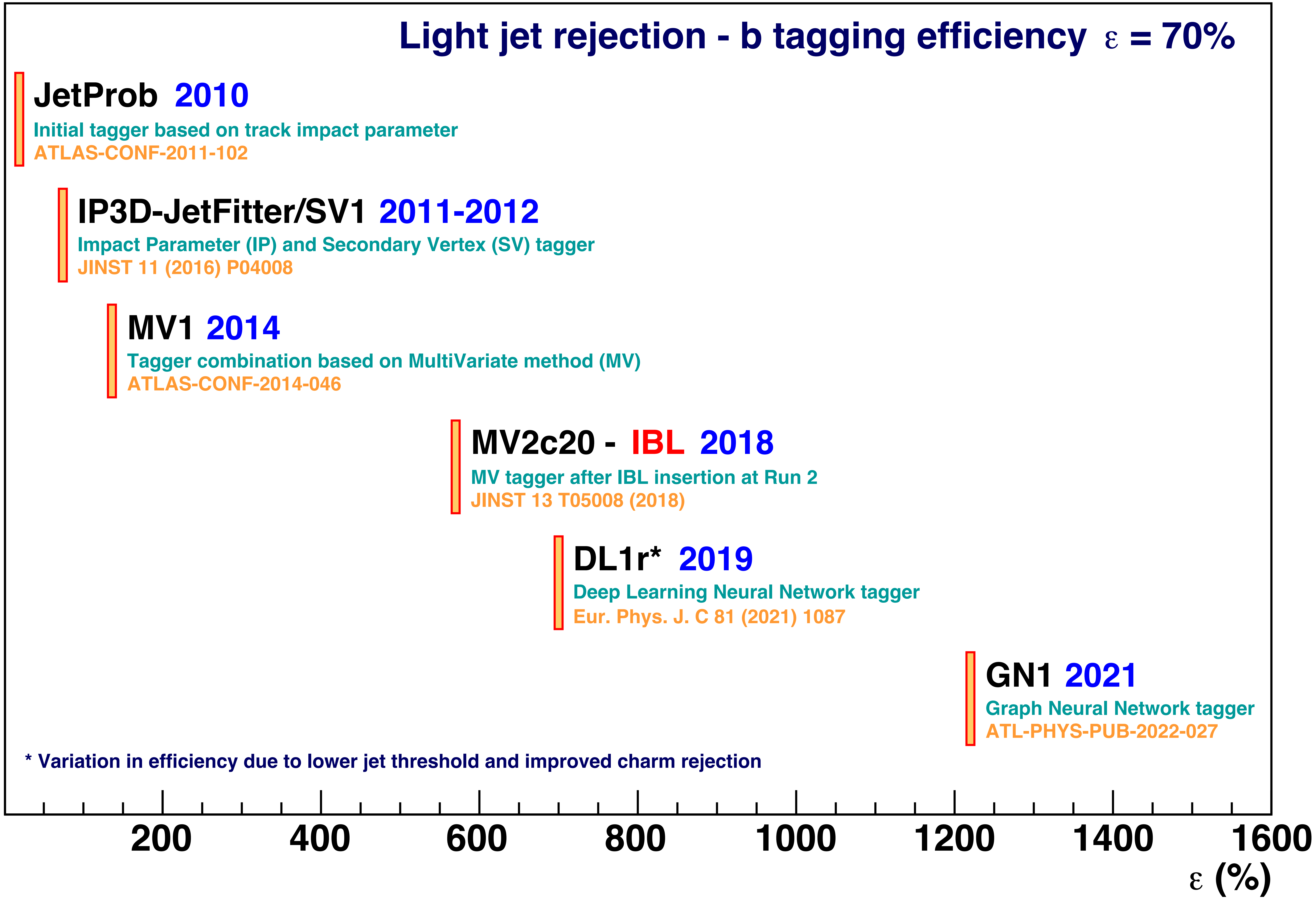
(same is true for charm)





the only reason we continue to see improvements in flavor-tagging ***in data***

is because we are able to model heavy-flavour production and decays to high accuracy.

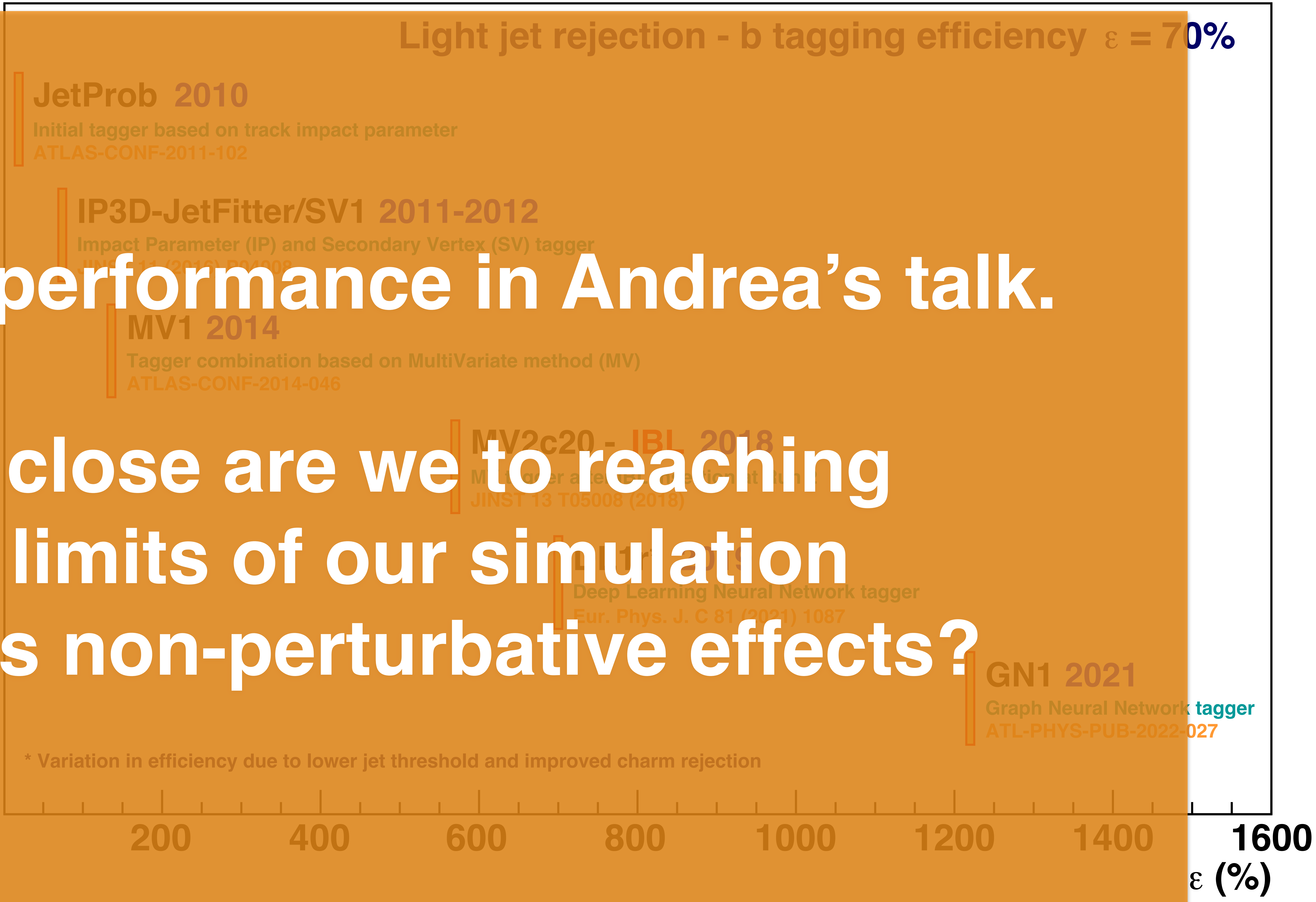


more on performance in Andrea's talk.

how close are we to reaching the limits of our simulation in terms non-perturbative effects?

the only reason we continue to see improvements in flavor-tagging ***in data***

is because we are able to model heavy-flavour production and decays to high accuracy.



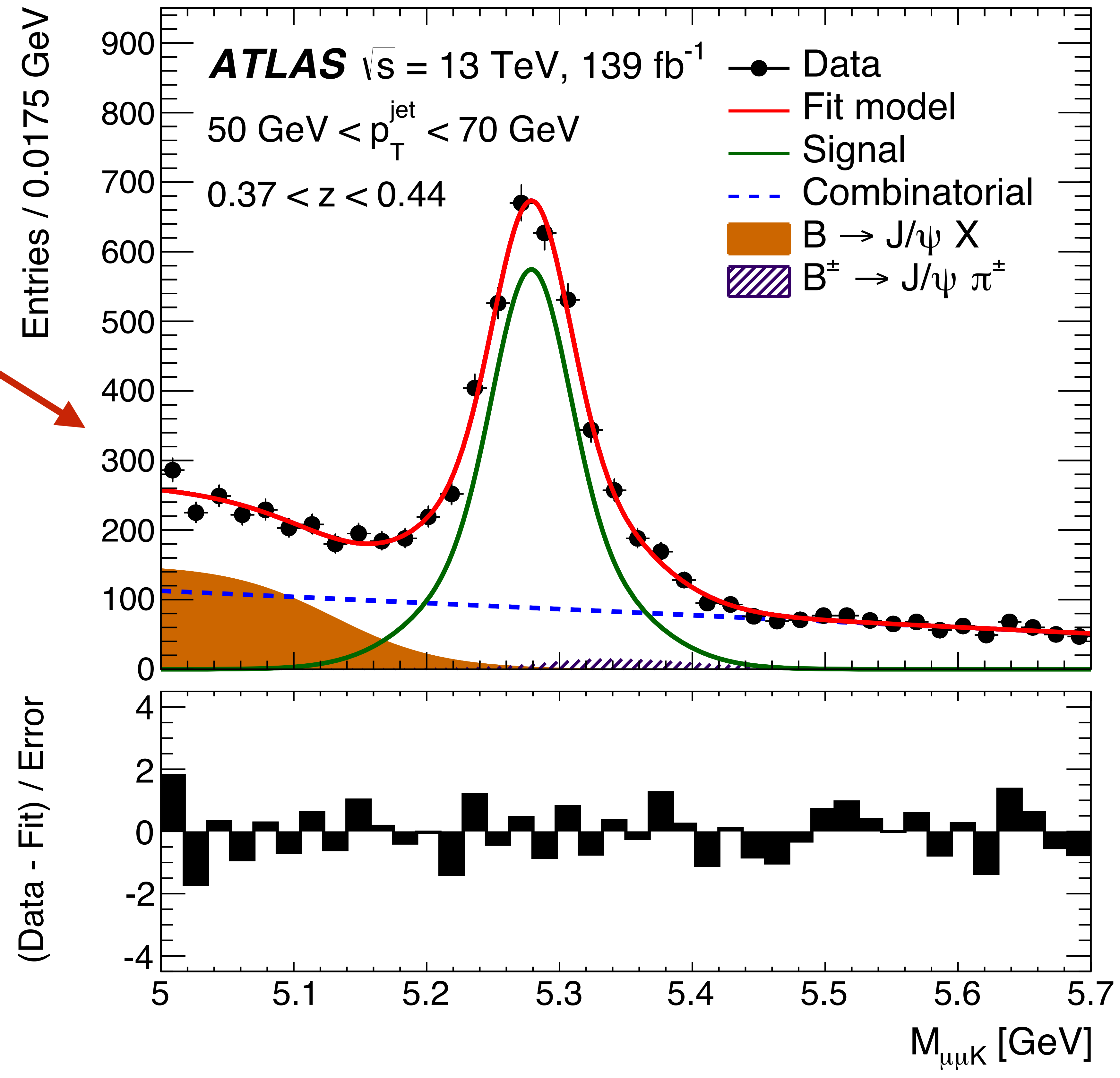
how?

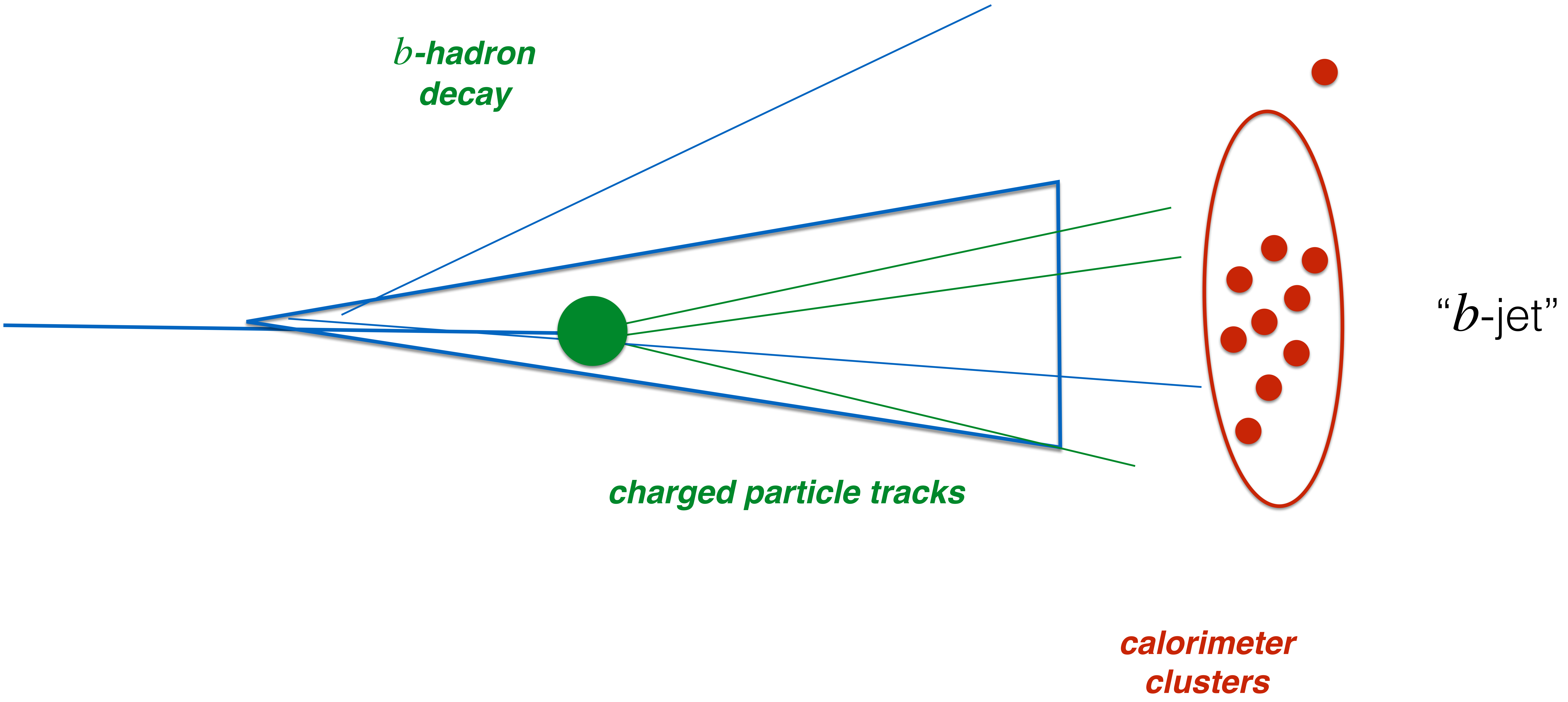
ATLAS recently released its first two measurements of b-quark fragmentation…

Short Title	Journal Reference	Date	√s (TeV)
b fragmentation in ttbar events at 13 TeV	Submitted to PRD	28-FEB-22	13
Exclusive b fragmentation at 13 TeV	JHEP 12 (2021) 131	26-AUG-21	13

- provide excellent coverage where LEP data can't reach
  - and *extremely complementary* to each other
- this is the "first generation" of such measurements
  - many aspects could be improved!

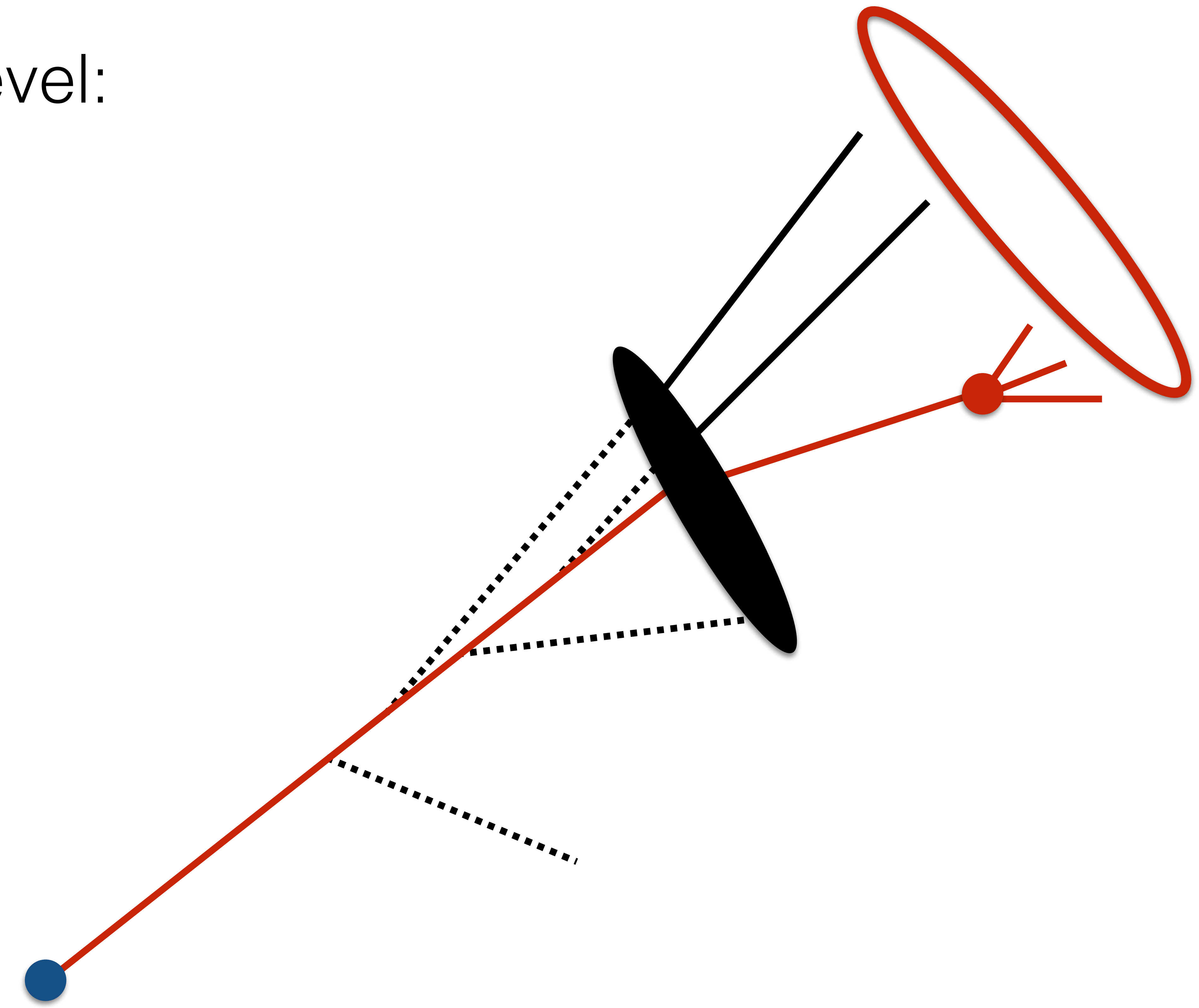
JHEP 12 (2021) 131, PRD 106 (2022) 032008

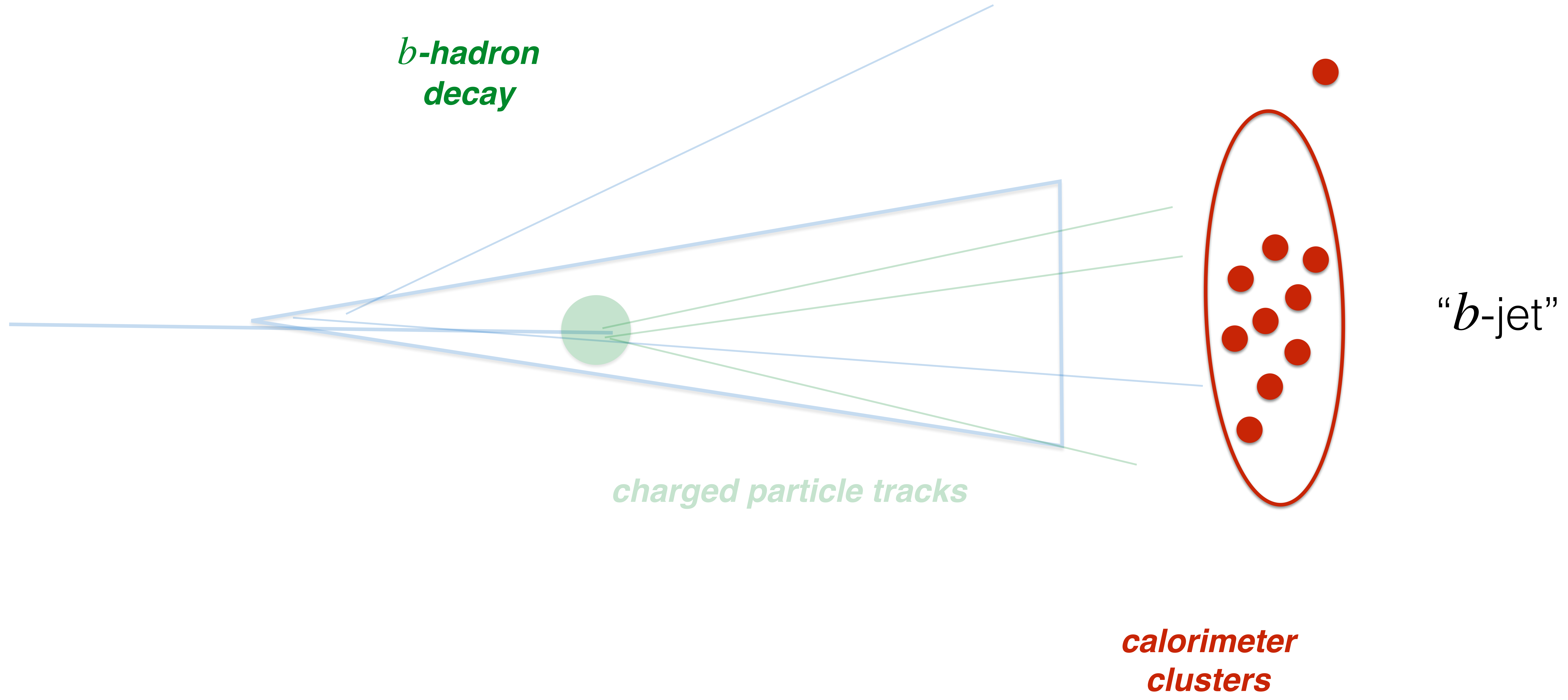




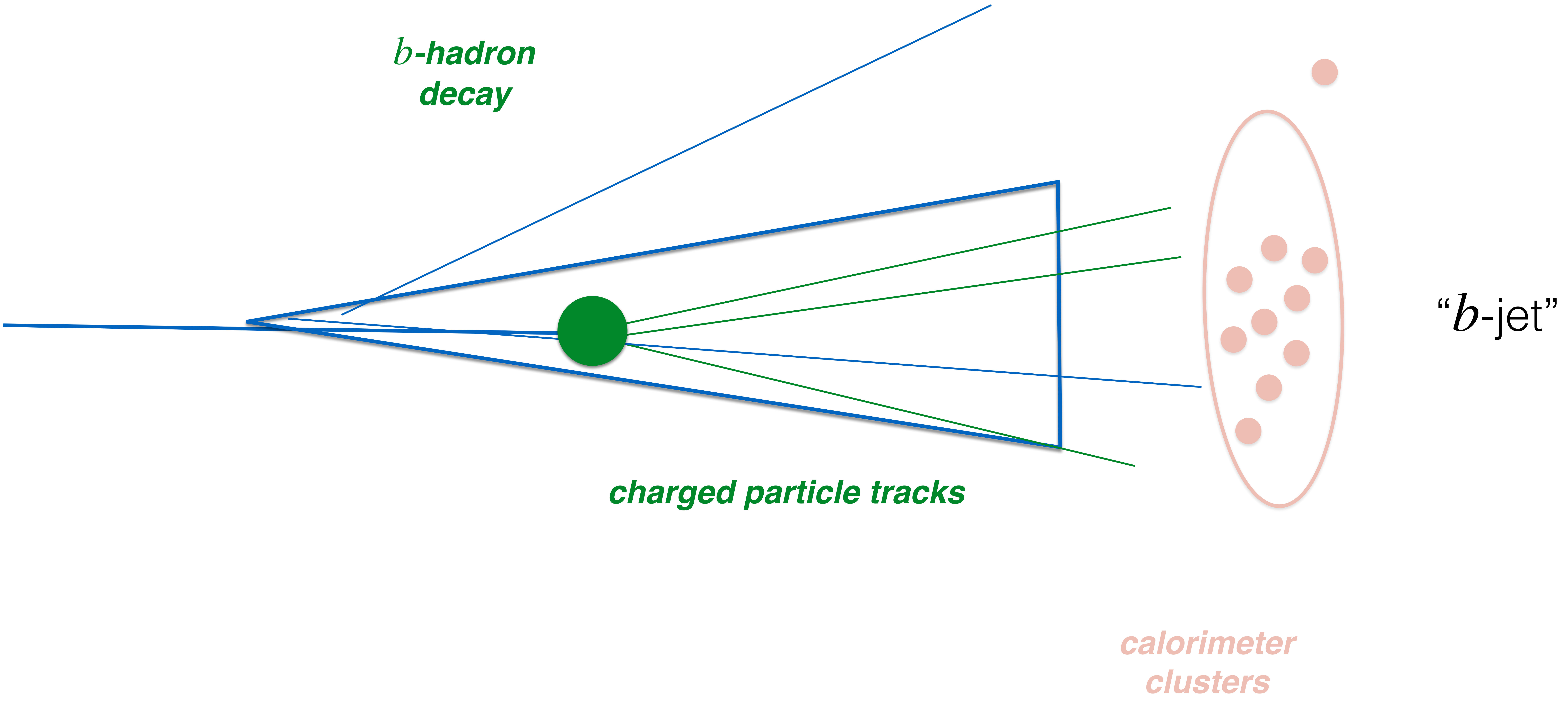
b-hadron decay
charged particle tracks
"b-jet"
calorimeter clusters

- both measurements unfold related observables to particle level:
  - $z_{(L)} = \vec{p}_B \cdot \vec{p}_{jet} / p_{jet}^2$
  - $p_T^{rel} = |\vec{p}_B \times \vec{p}_{jet}| / |p_{jet}|$ ($B \to \mu\mu K$ only)
  - $\rho = p_T^B / \mathrm{avg}(p_T^{\ell})$ ($t\bar{t}$ only)
  - charged particle multiplicity, n_{ch}^B ($t\bar{t}$ only)
- multijet: measure full $B \to \mu\mu K$ and full jet momentum
- $t\bar{t}$: measure "charged momentum" of b-hadron and jet

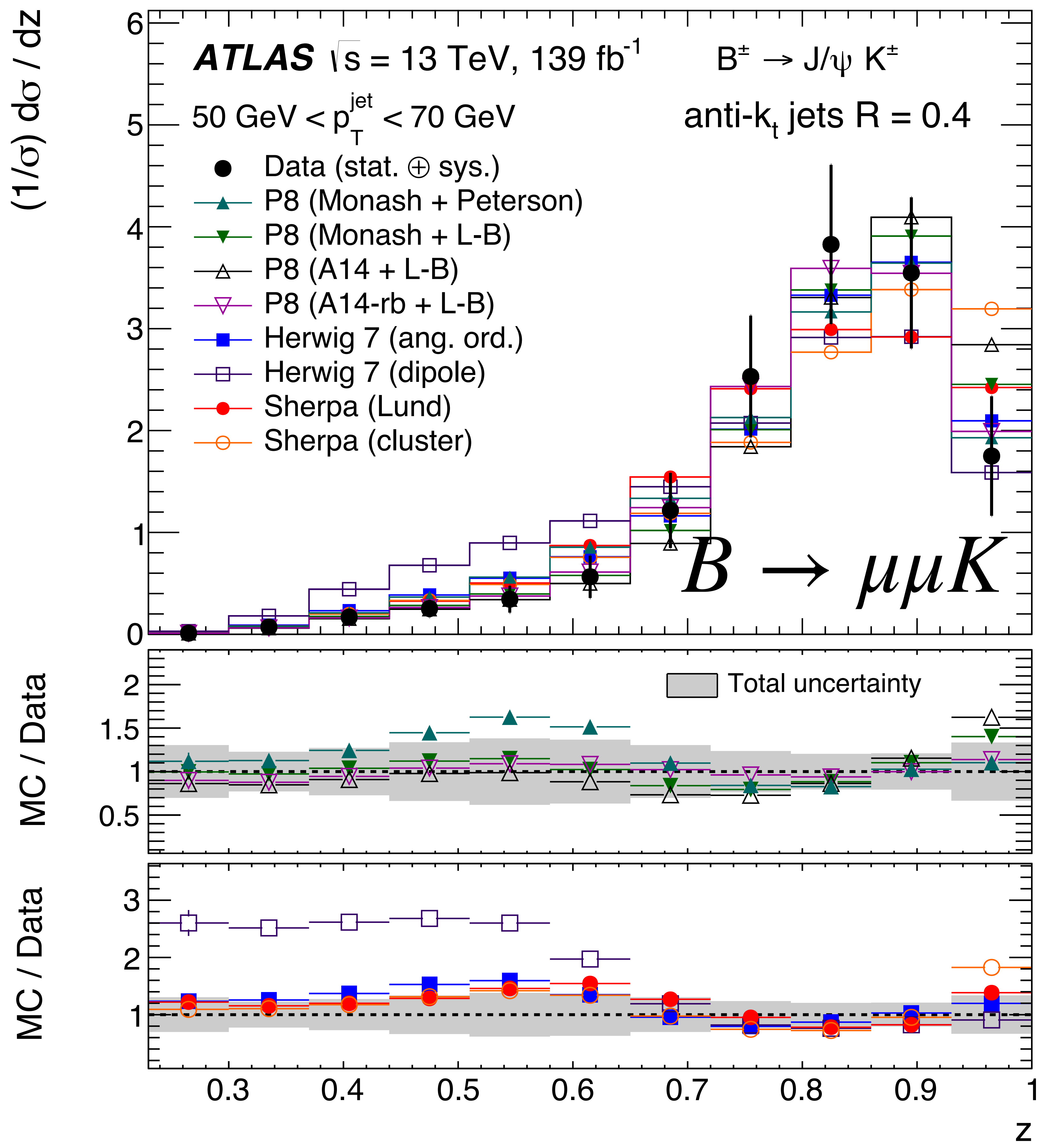


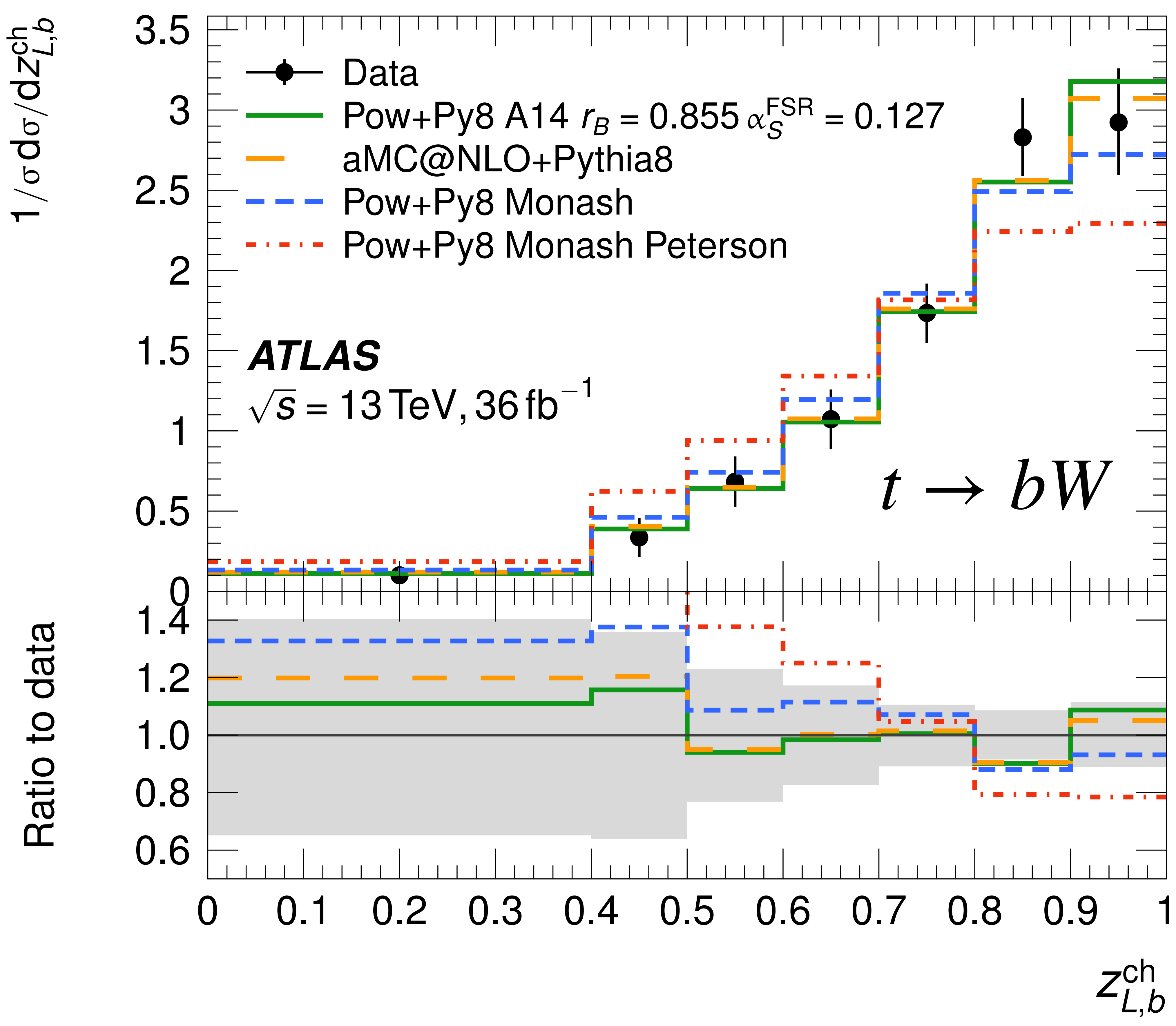


multijet: measure full $B \to \mu\mu K$ and full jet momentum



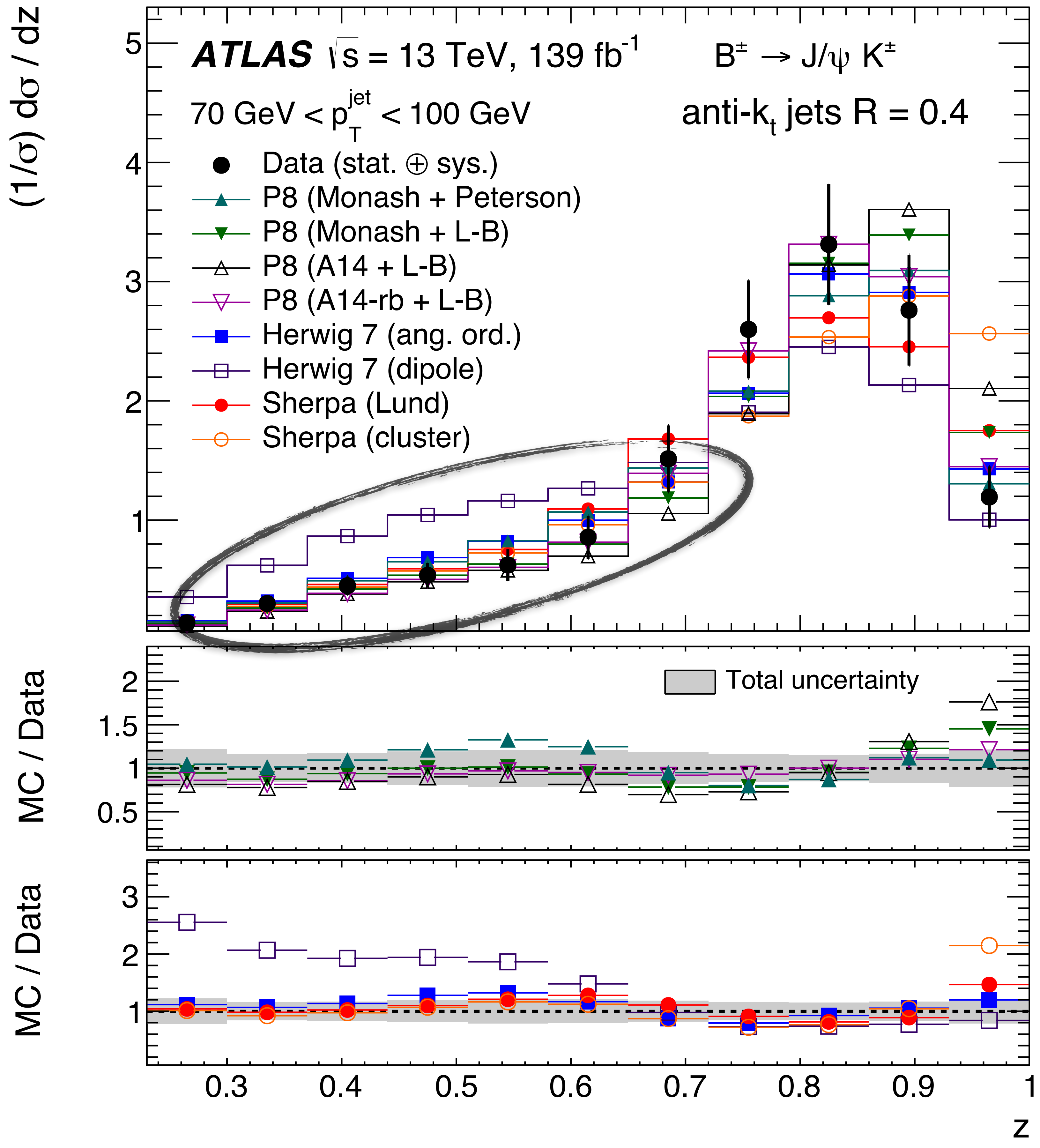
$t\bar{t}$: measure "charged momentum" of b-hadron and jet





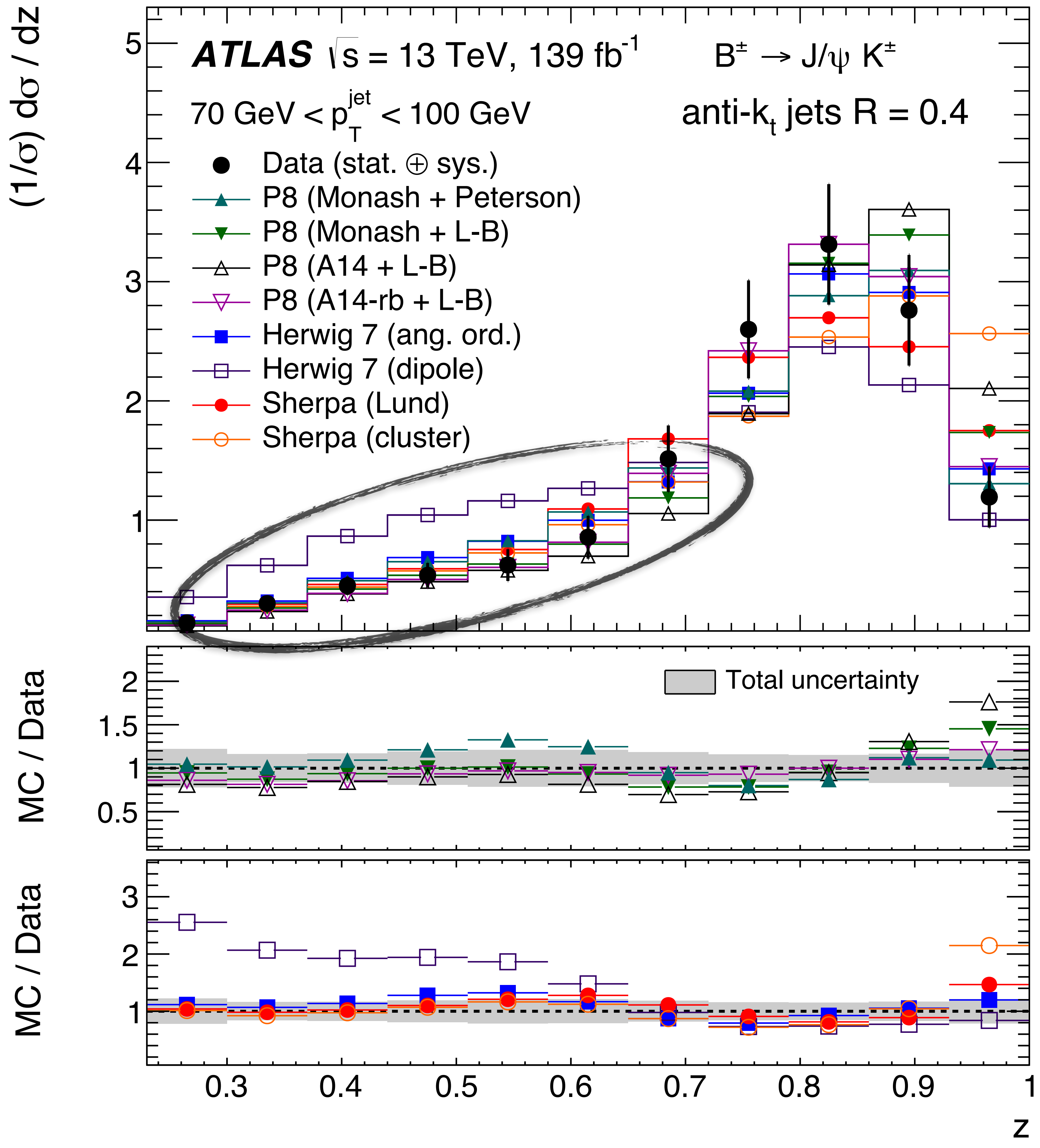
- calculations shown are **NLO+PS+tune.**
- good discriminating power between leading MC generators/tunes.

$B \to \mu\mu K$



clear issues with low-z spectrum for some calculations

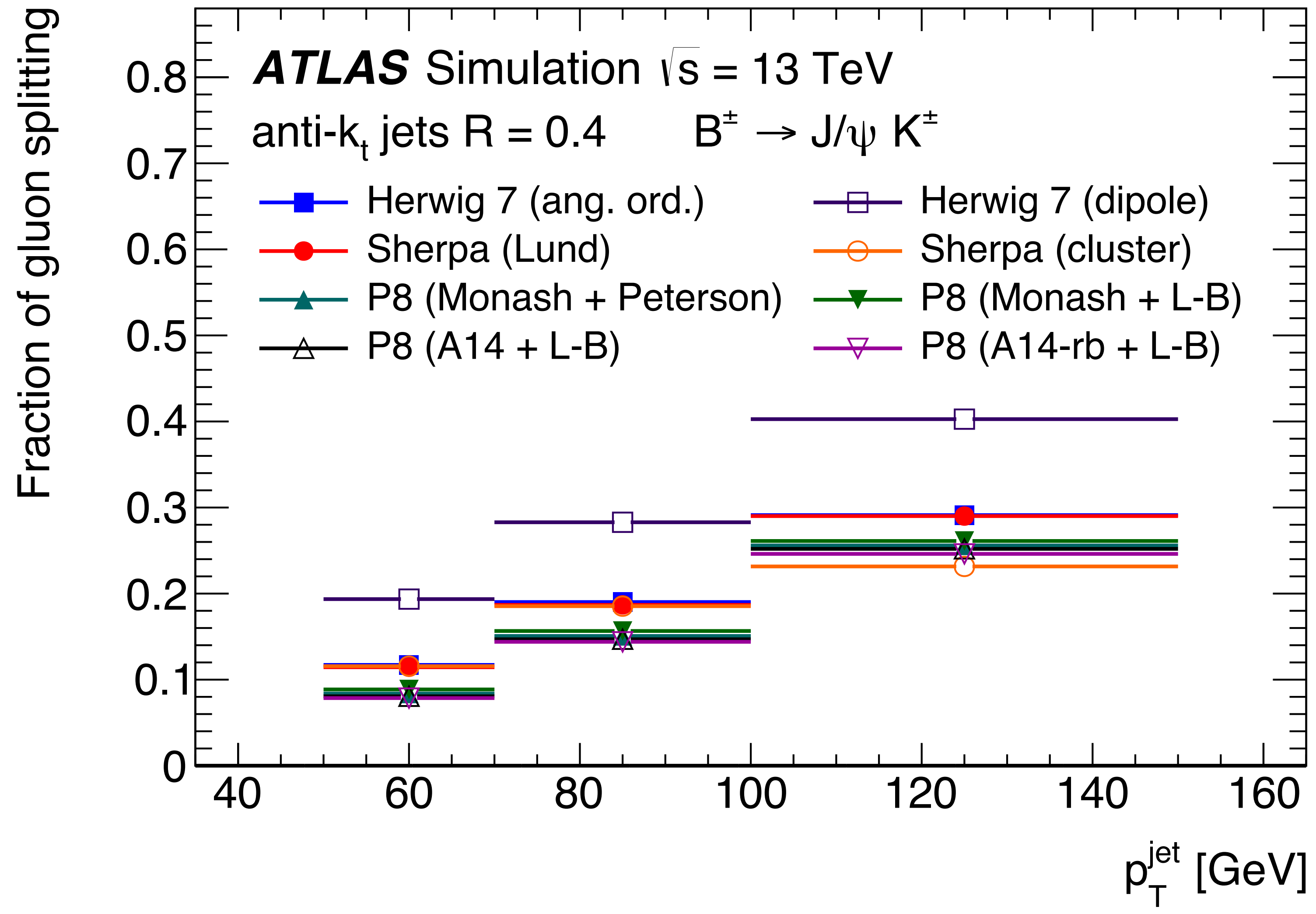
$B \rightarrow \mu\mu K$

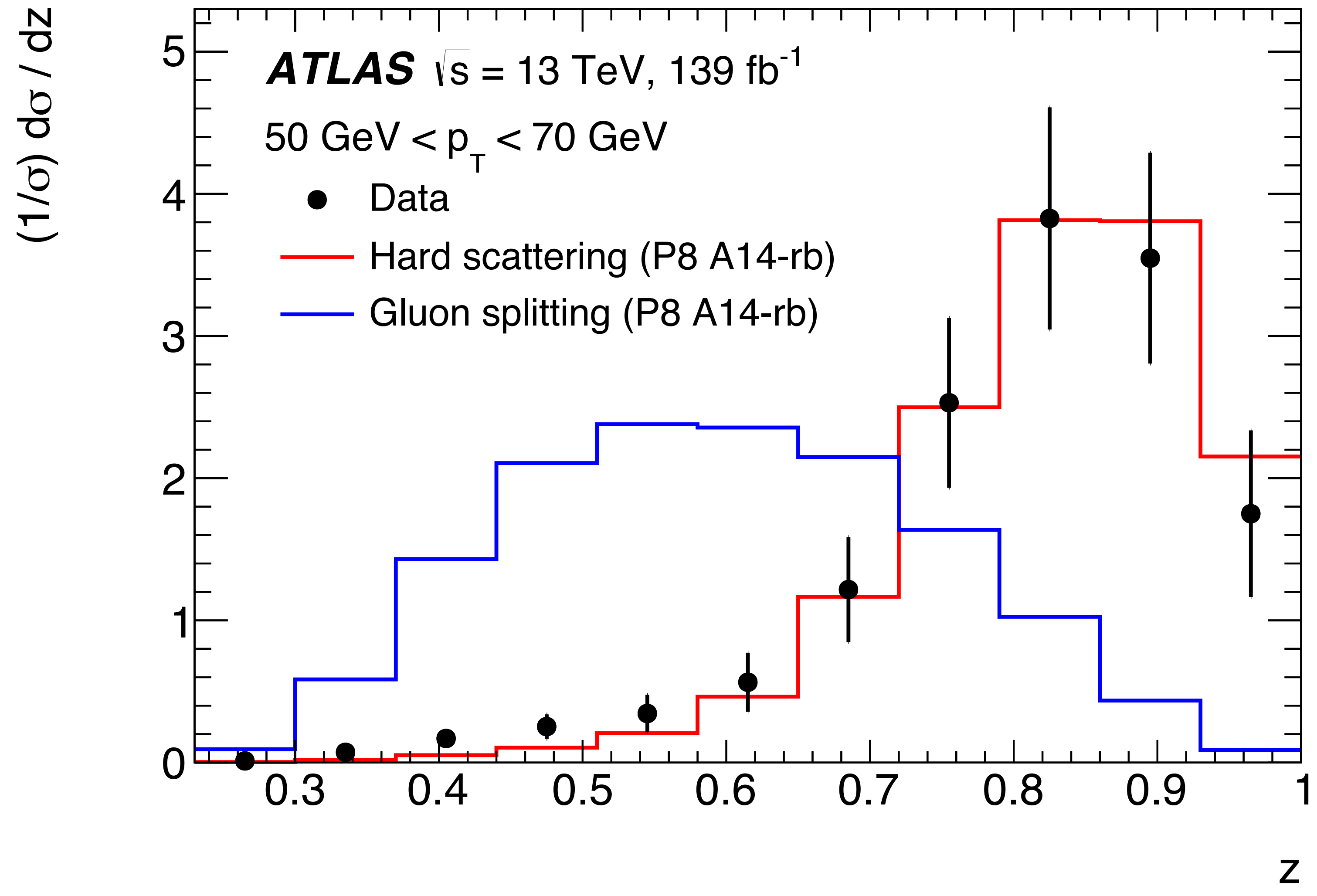


clear issues with low-z spectrum for some calculations

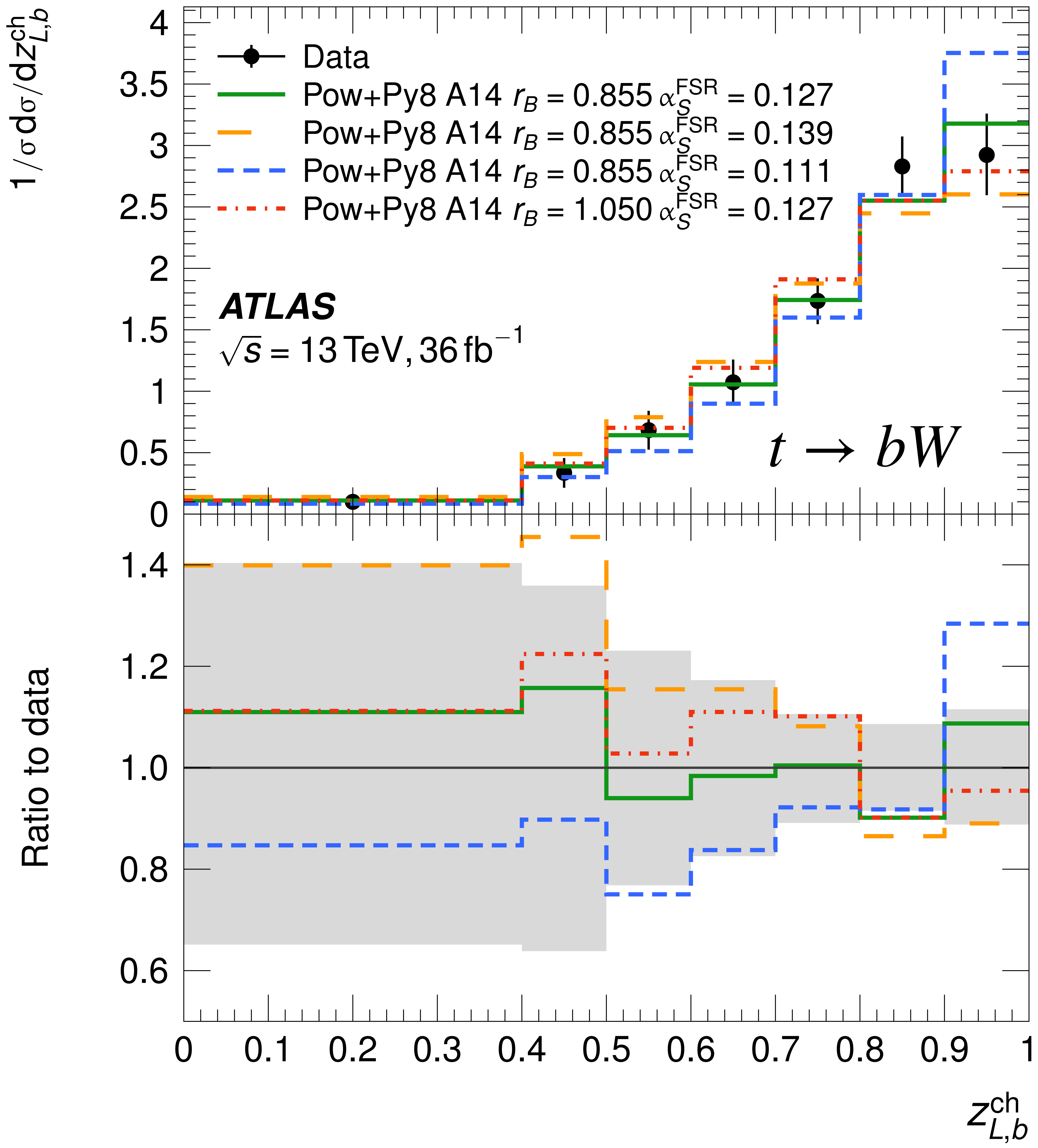
likely due to mismodeled $g \rightarrow bb$ rates

$t \rightarrow bW$ analysis can disentangle effects (no $g \rightarrow bb$ jets)

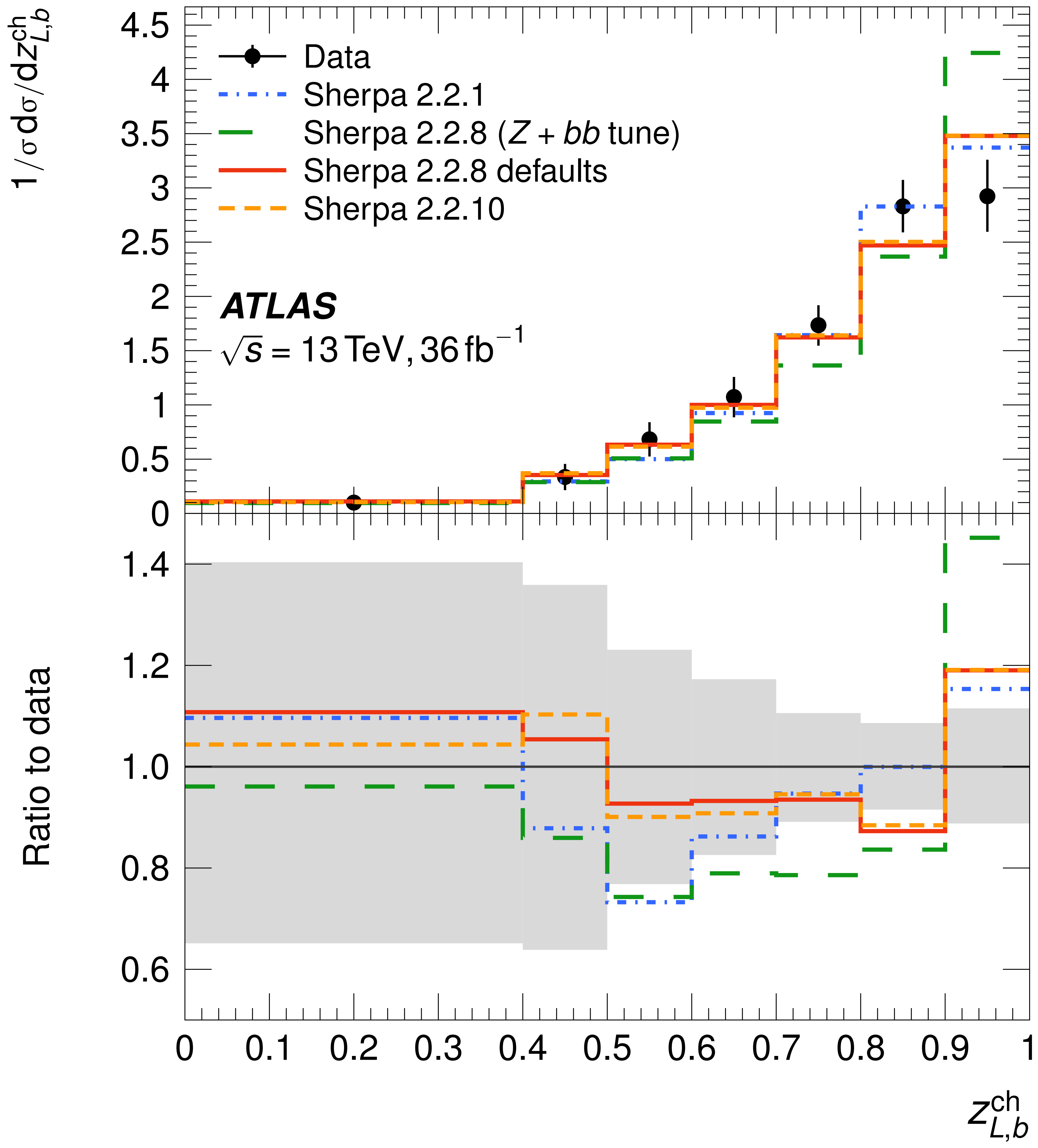




Data
Pow+Py8 A14 $r_B = 0.855$ $\alpha_S^{FSR} = 0.127$
Pow+Py8 A14 $r_B = 0.855$ $\alpha_S^{FSR} = 0.139$
Pow+Py8 A14 $r_B = 0.855$ $\alpha_S^{FSR} = 0.111$
Pow+Py8 A14 $r_B = 1.050$ $\alpha_S^{FSR} = 0.127$

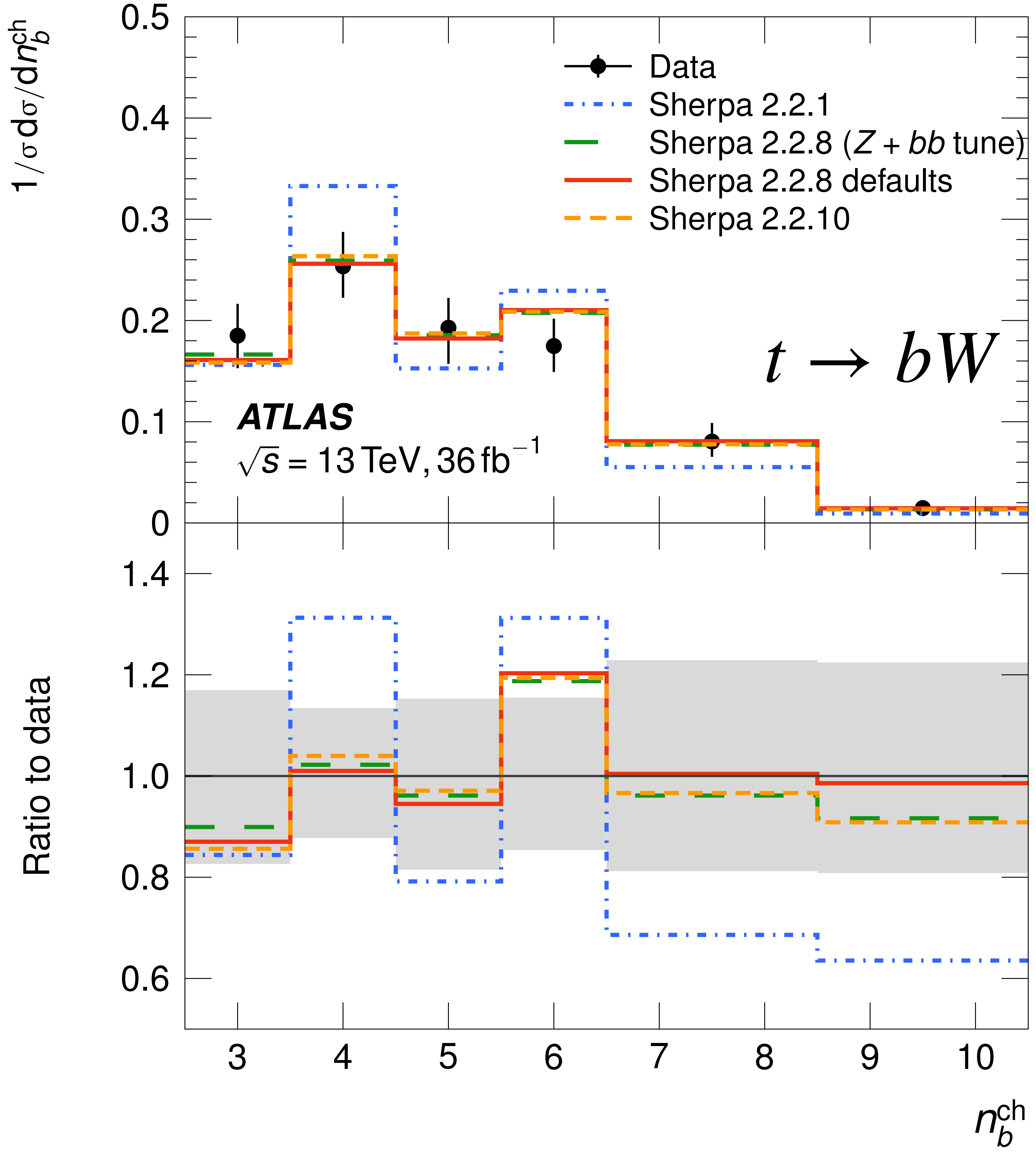


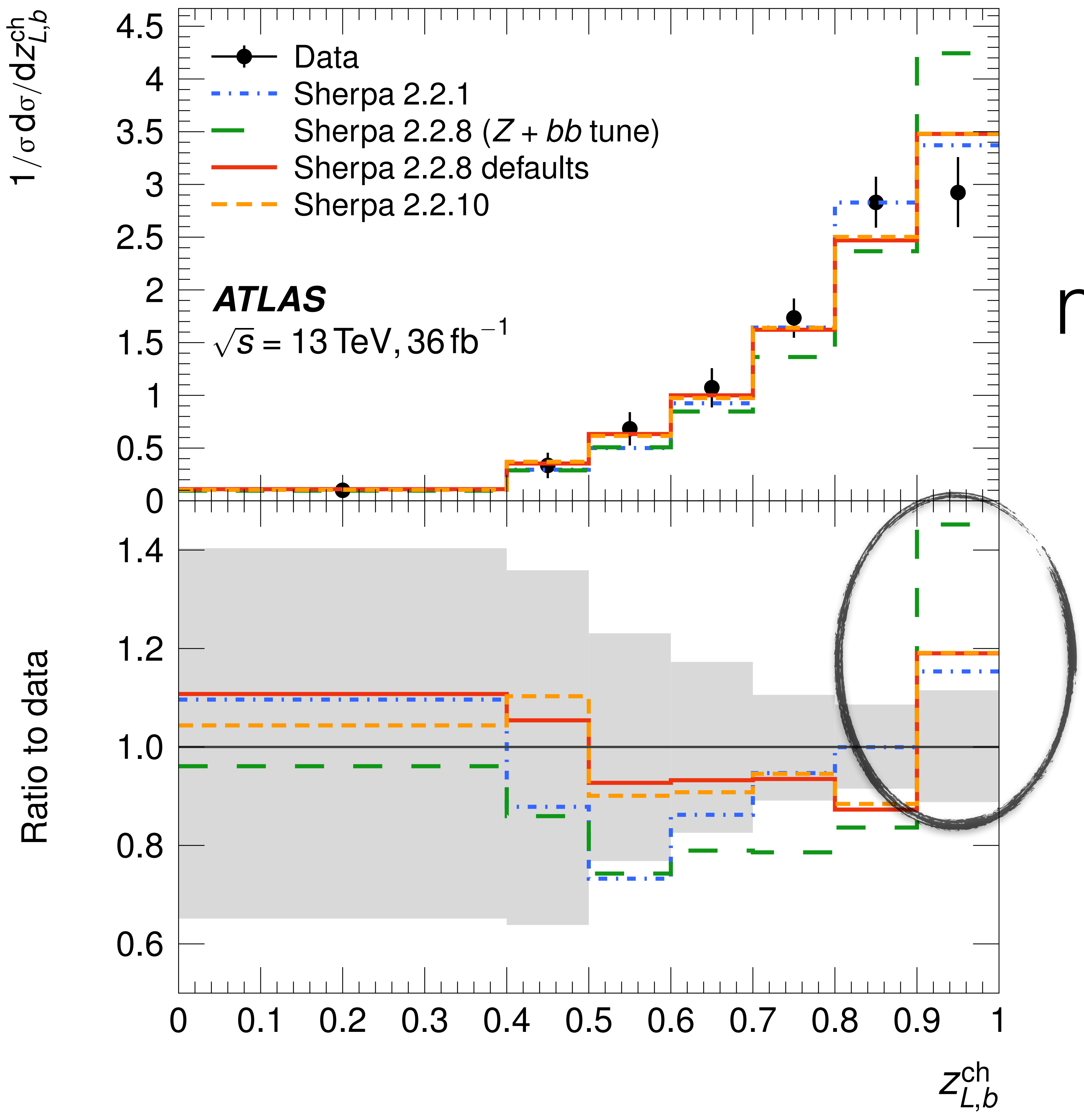
clear correlation
between choice of α_S^{FSR}
and fragmentation parameters (r_B)
needs to be considered.



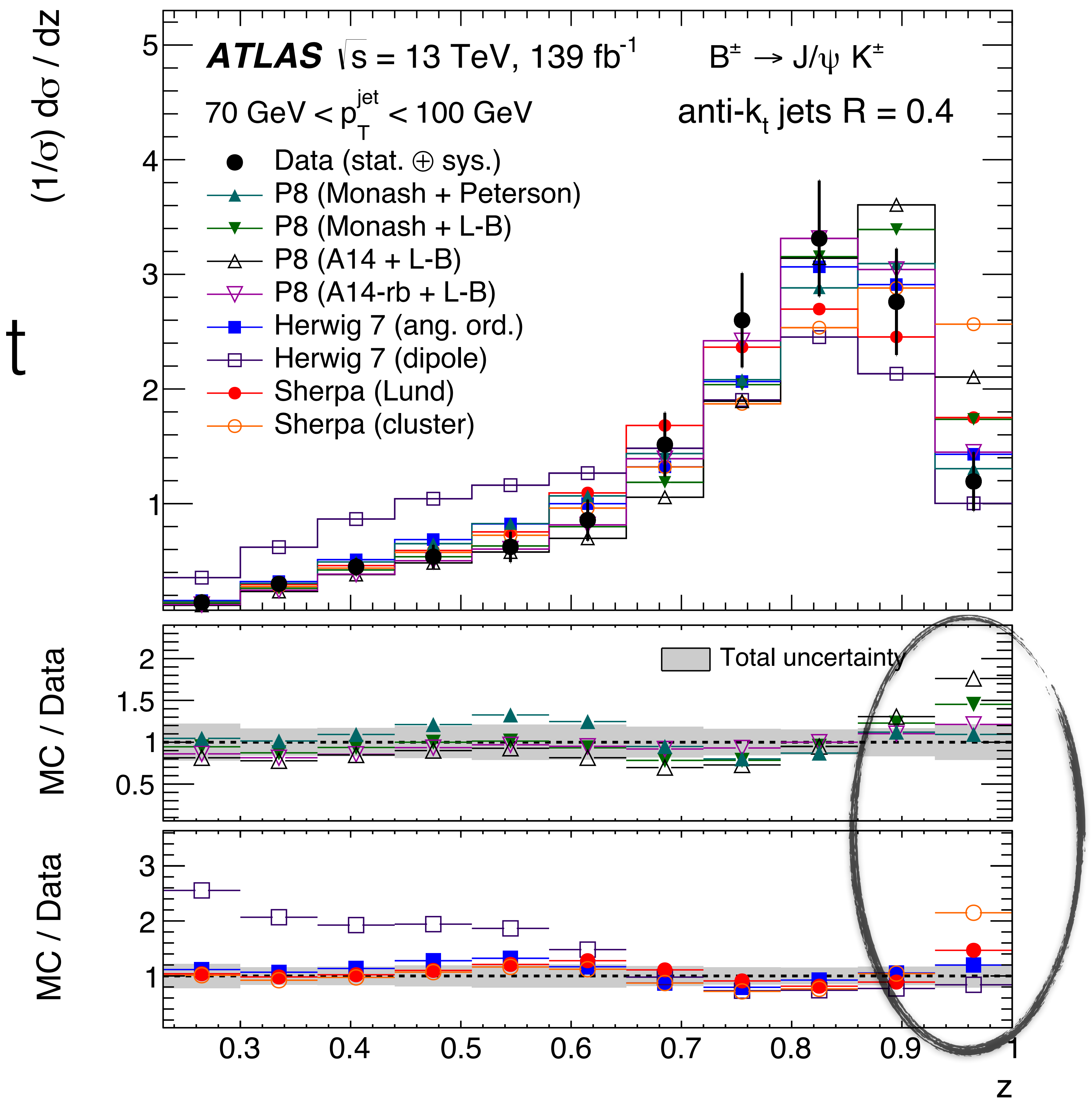
beware!

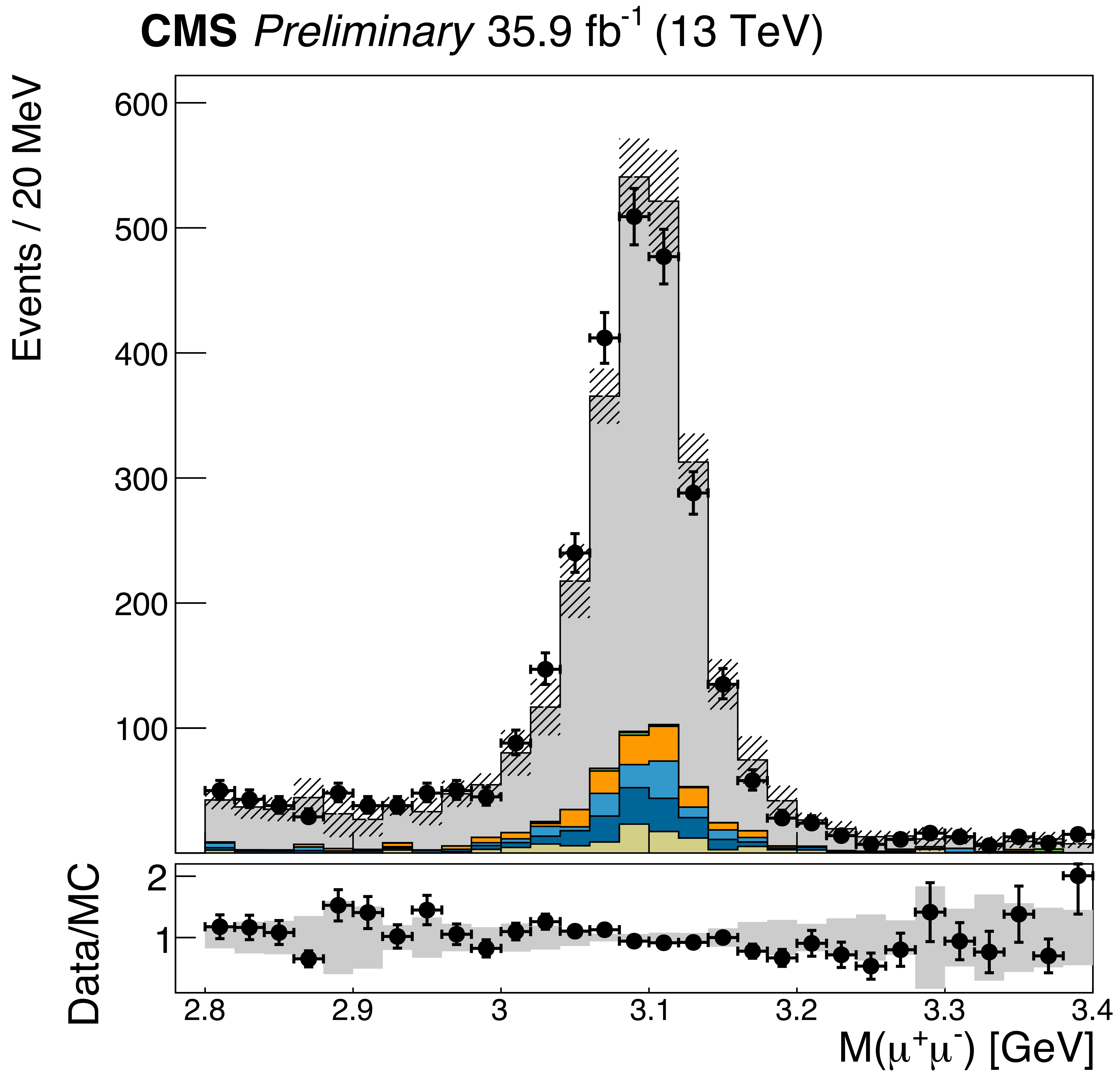
details of b-fragmentation and top-quark decays *depend strongly on parameters that need to be carefully validated in MC generators*.





correlated
mismodeling independent
of b-quark source





CMS PAS TOP-18-012

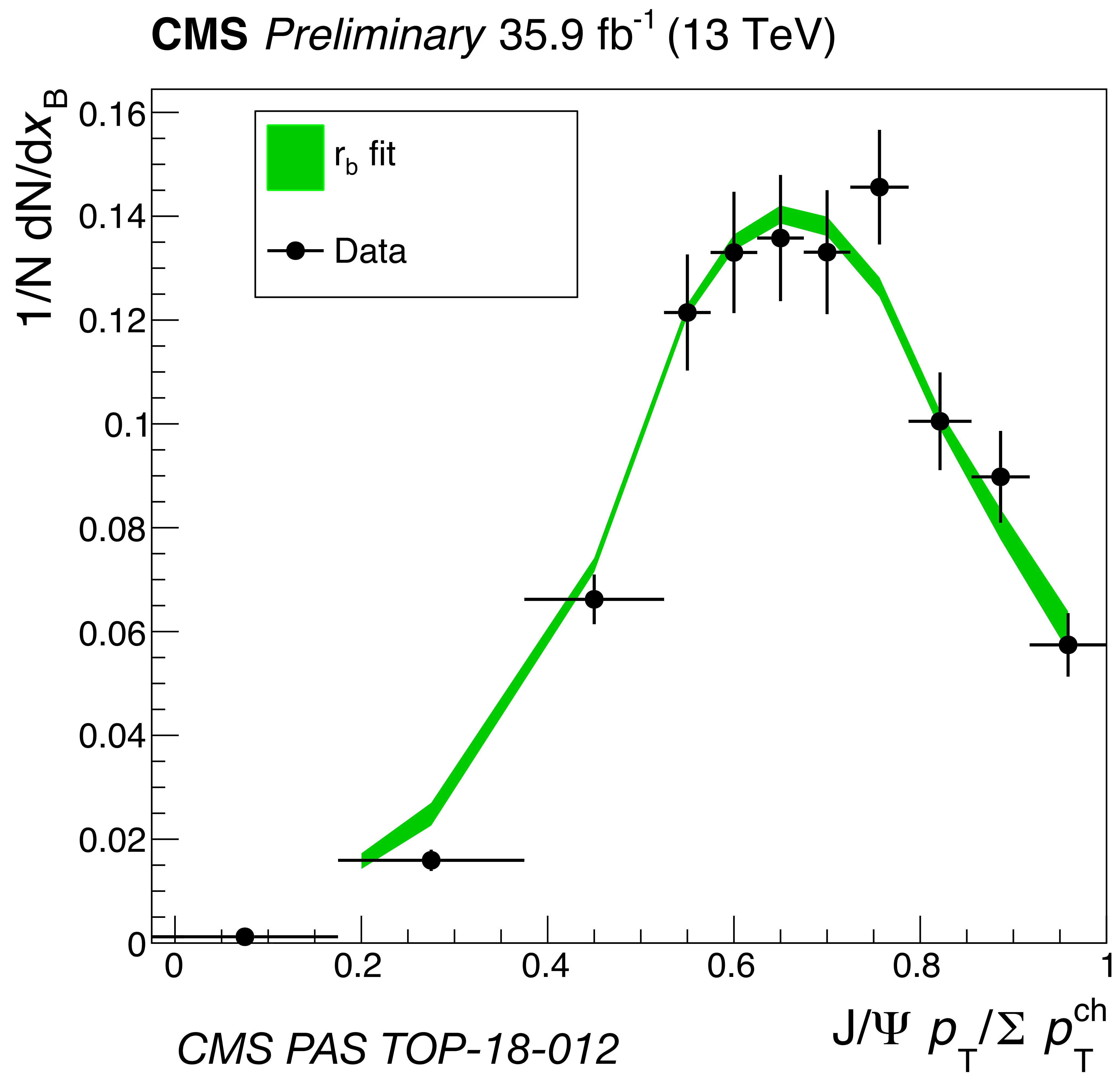
CMS have extracted the (pythia) Lund-Bowler parameter r_b

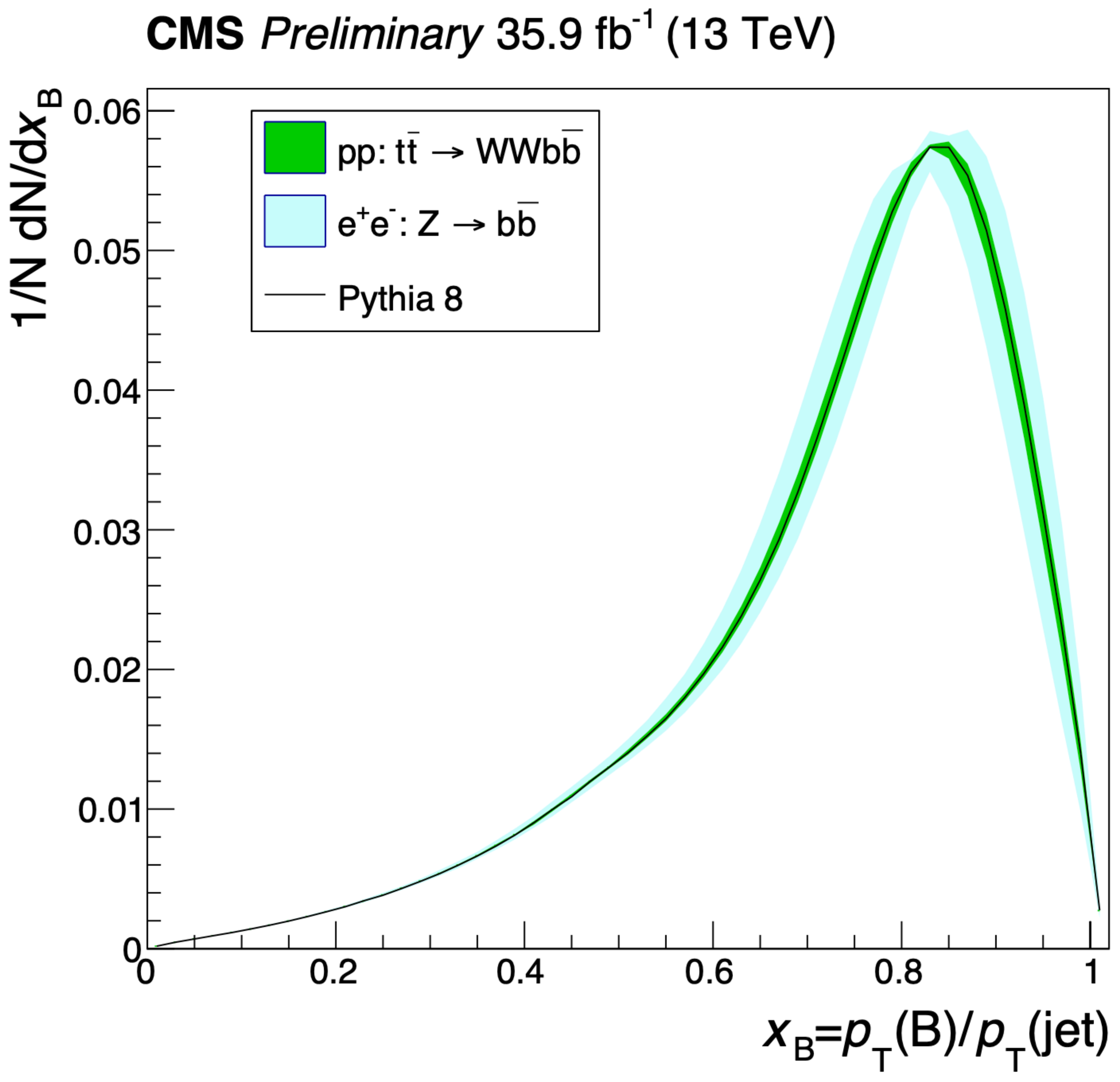
$$f(z) = \frac{1}{z^{1+r_q b m_q^2}} (1-z)^a \exp\left(-\frac{b m_T^2}{z}\right)$$

uses fully-reconstructed $D^0 \to K^\pm \pi^\mp$ and $J/\psi \to \mu\mu$ from b-hadron decays

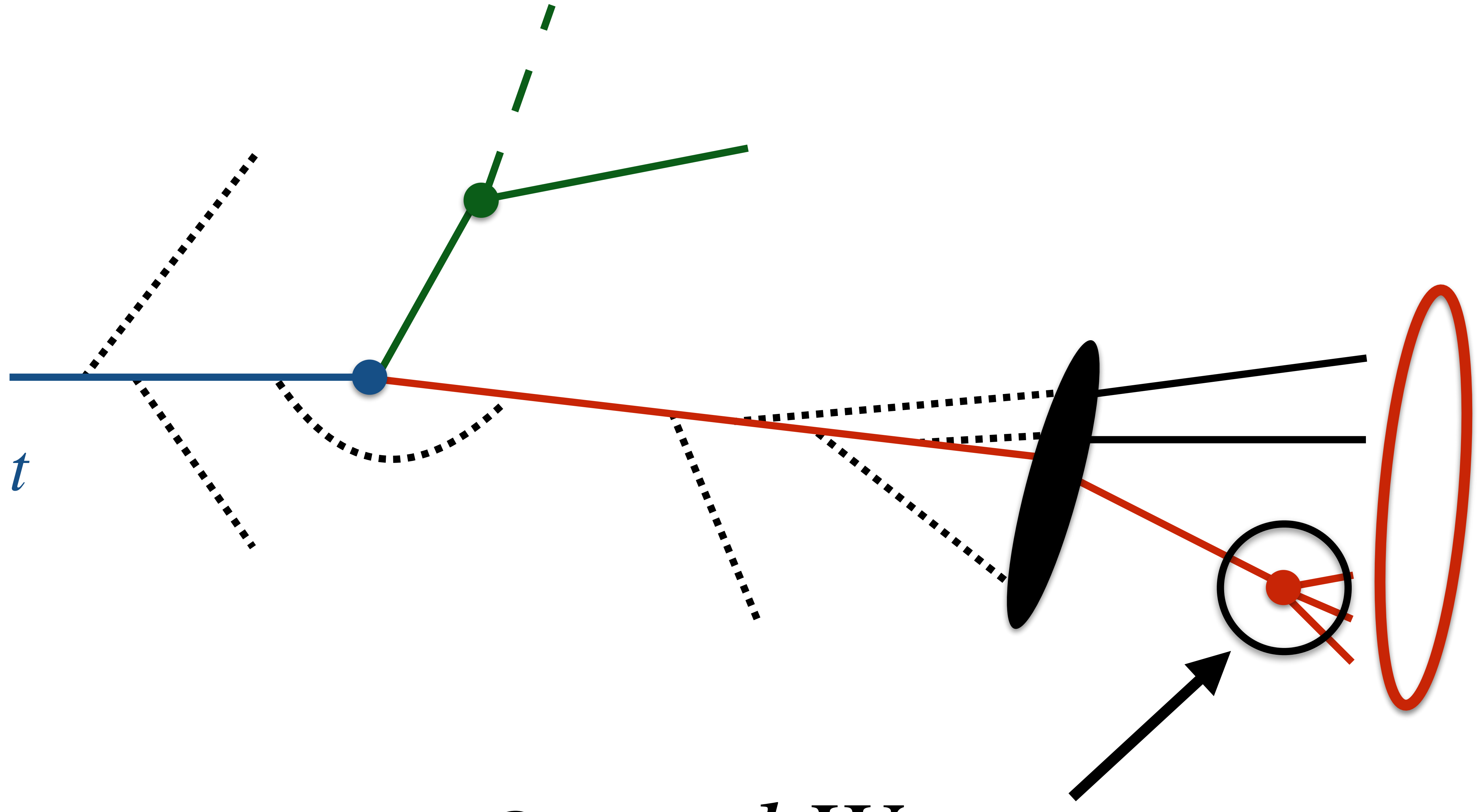
direct, detector-level fit to r_b

simultaneous fit to r_b across channels: $r_{\mathrm{b}} = 0.858 \pm 0.037\,(\mathrm{stat}) \pm 0.031\,(\mathrm{syst})$





where to?

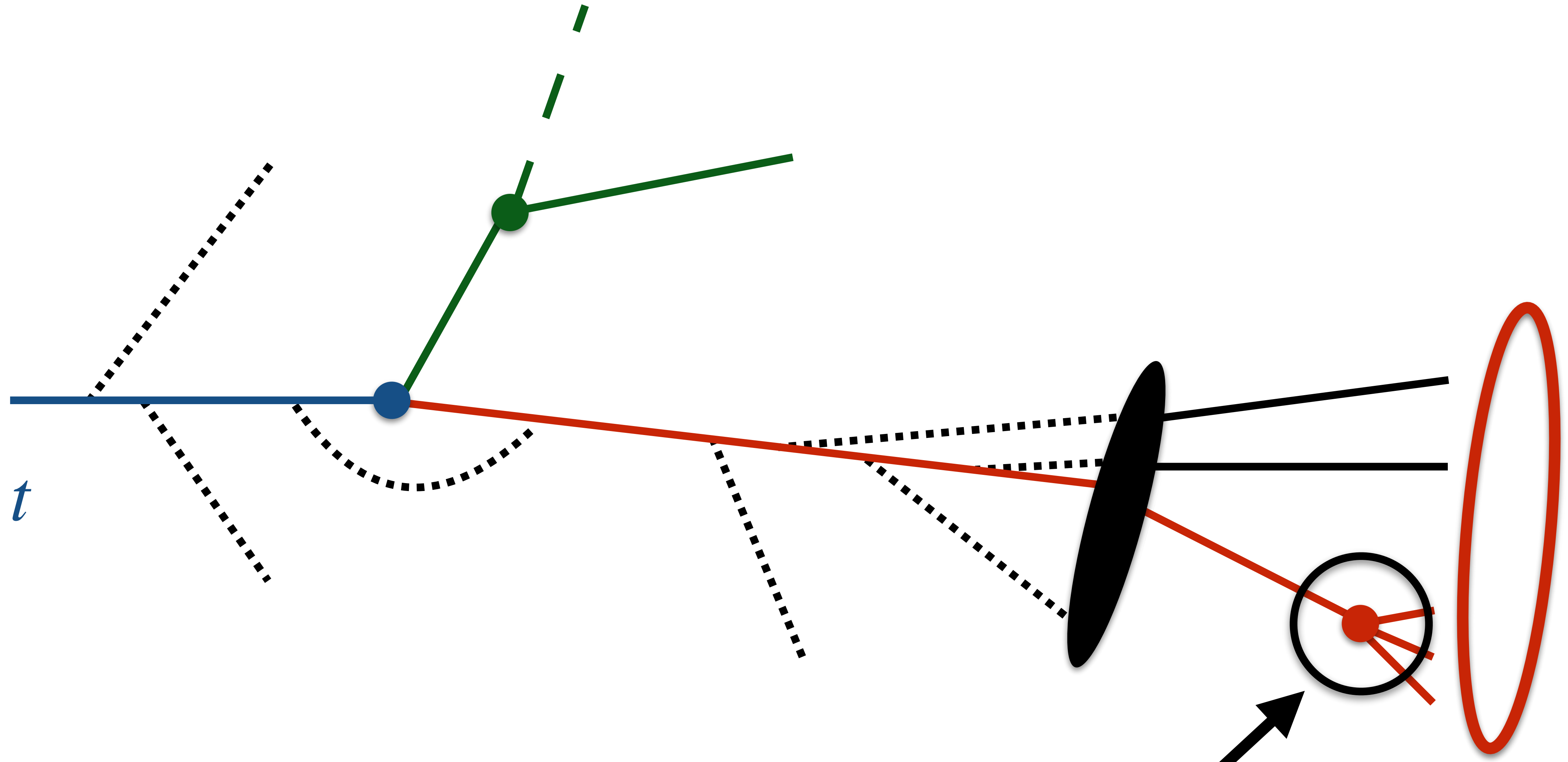


ATLAS $t \to bW$ measurements are limited by the resolution of the b-hadron decay vertex.

there are a few ways to improve here:

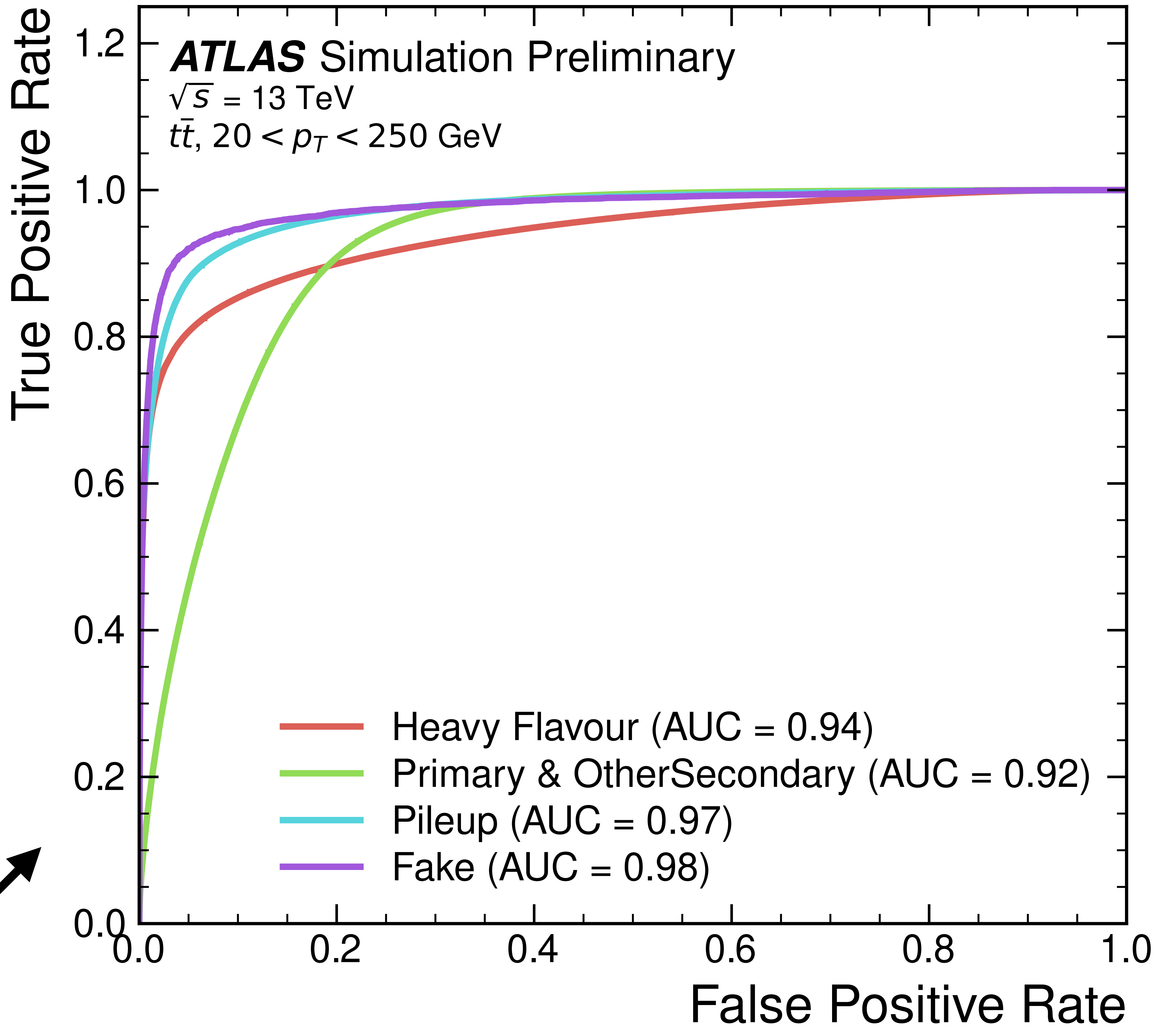
1) use more exclusive decay modes (requires more data)

2) derive better secondary-vertex algorithms



Graph Neural Network-based reconstruction is several factors better than conventional methods.

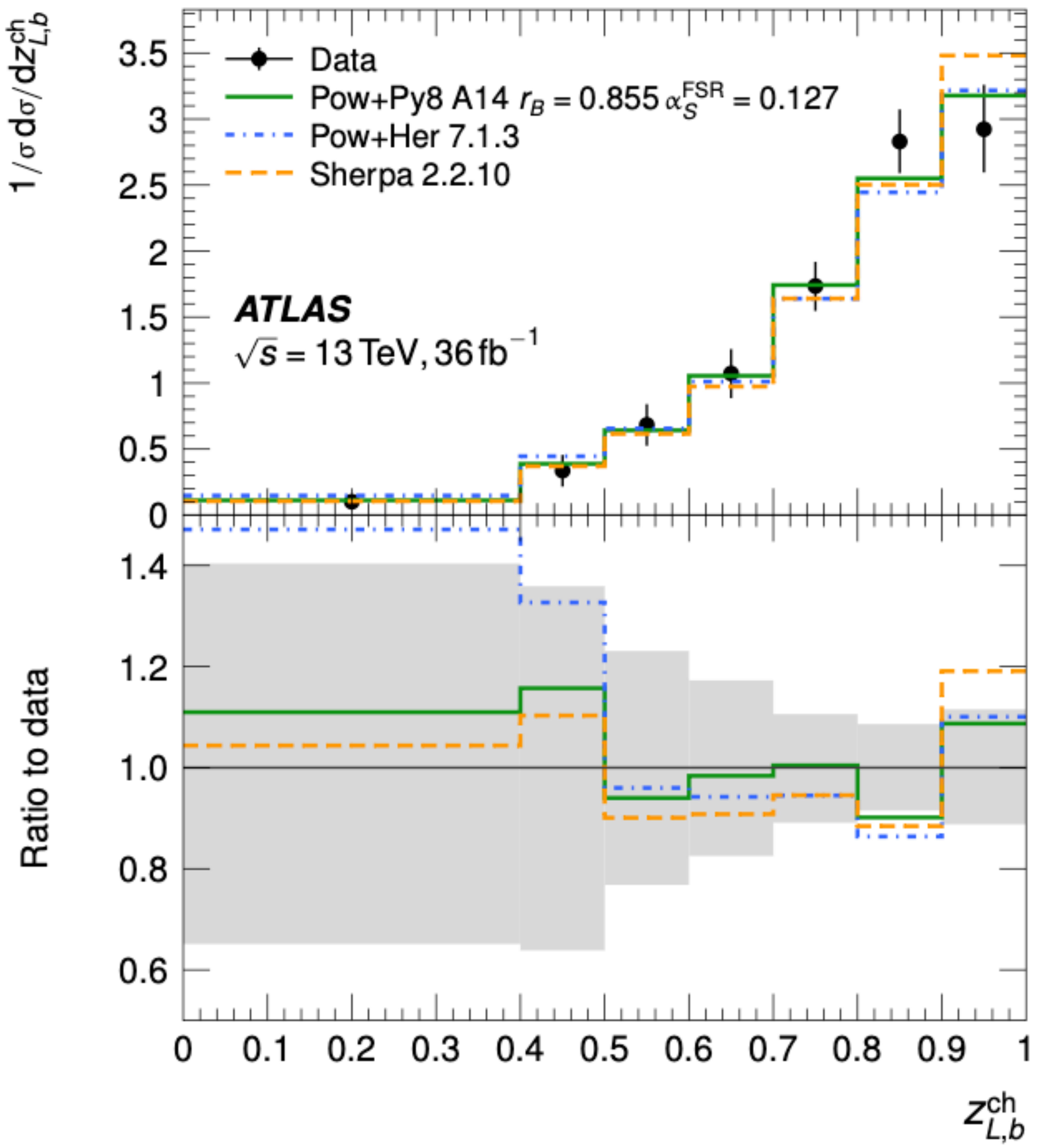
ATLAS $t \to bW$ measurements are limited by the resolution of the b-hadron decay vertex.

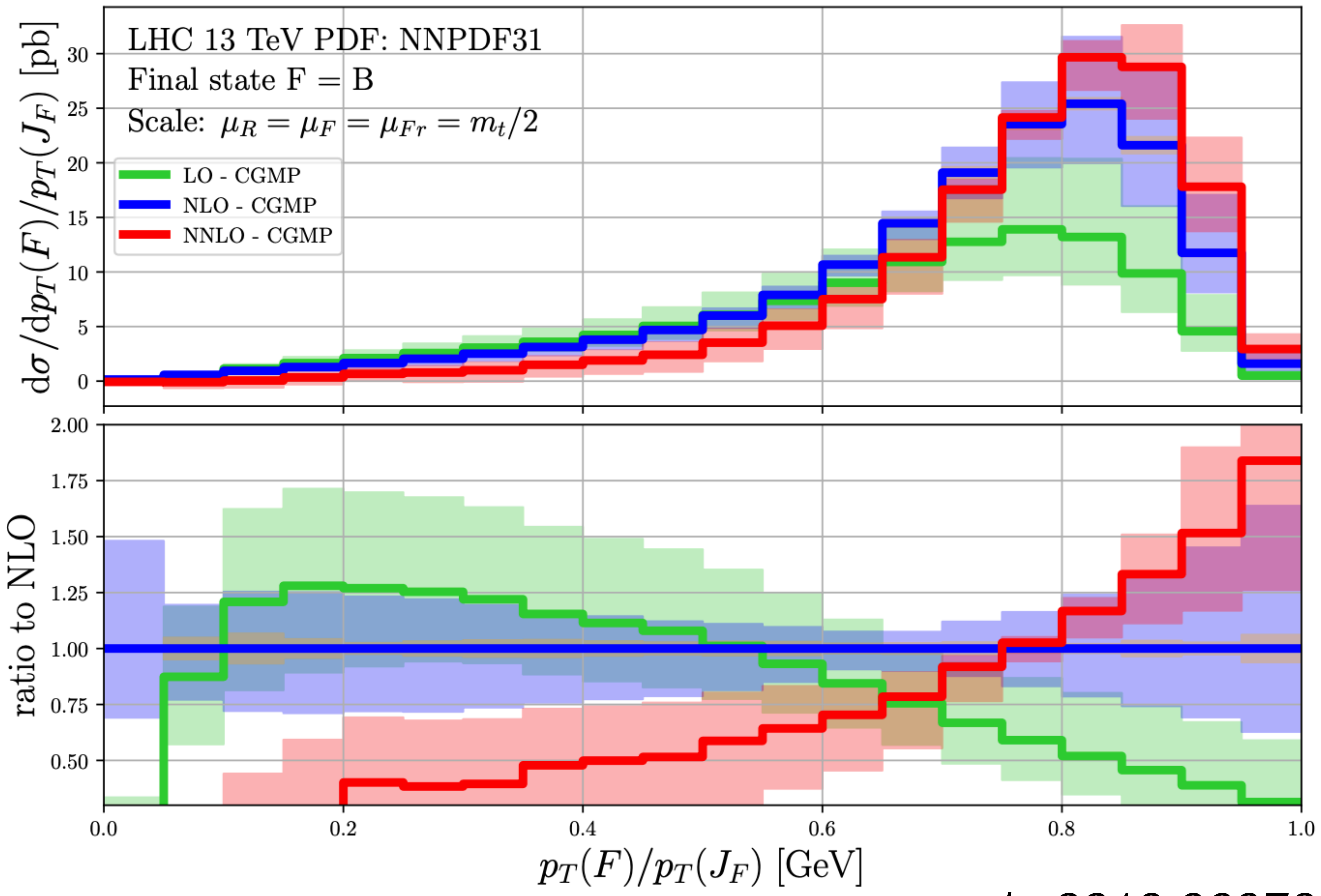


there are a few ways to improve here:

1) use more exclusive decay modes (requires more data)

2) derive better secondary-vertex algorithms





arxiv:2210.06078

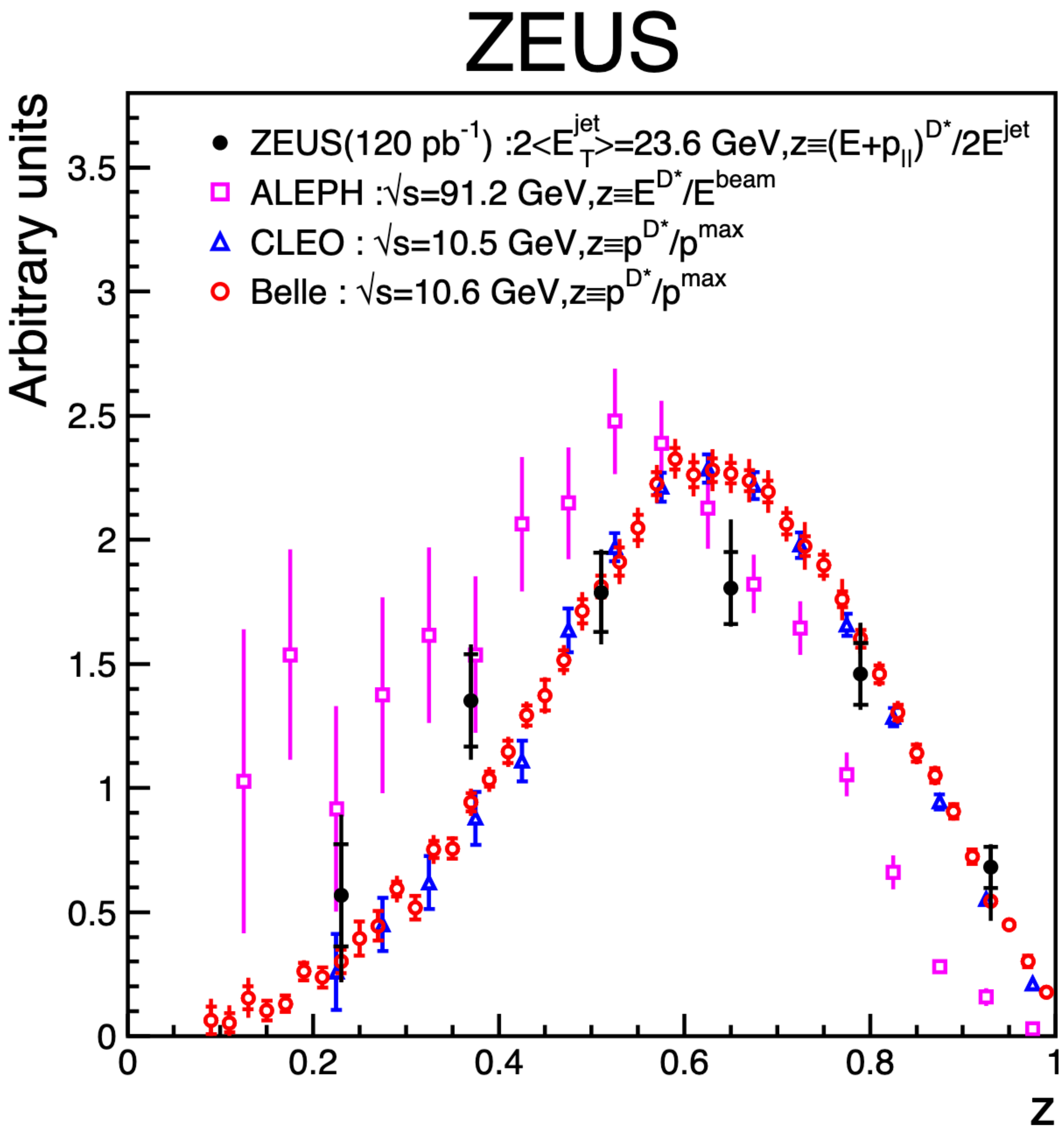
track-based measurements
are experimentally clean

but we need to work a bit to make
comparisons between unfolded
data and fixed-order predictions possible.

we need charm fragmentation measurements to support $h \to cc$ and strange PDF constraints.

we have a wealth of $W + c$ and $t \to cqb$ LHC data already being used to calibrate flavor-tagging algorithms → we should be measuring here!

($Z \to cc$ measurements from e^+e^- difficult to interpret…)

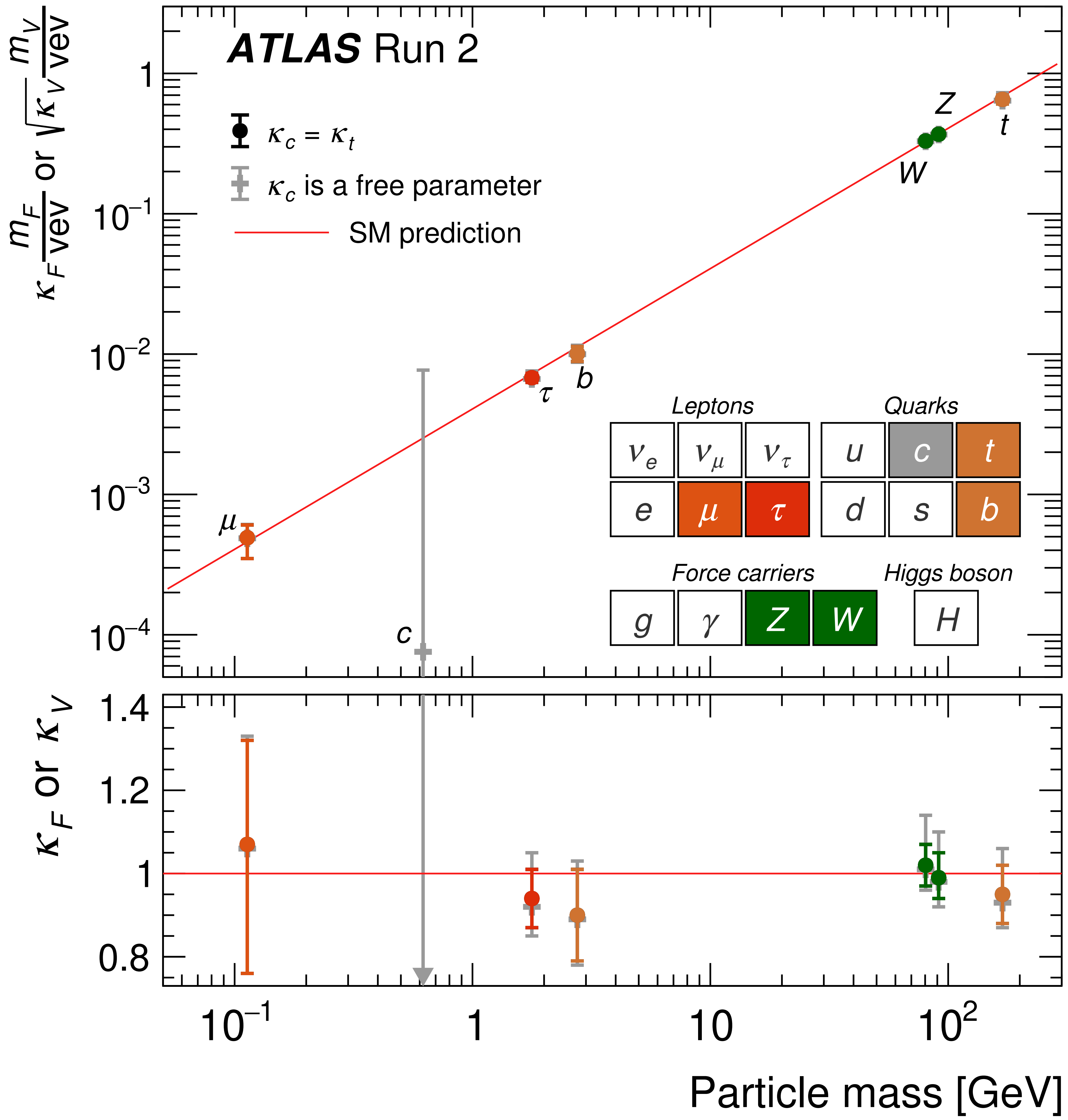


precision physics in the top, Higgs, and electroweak sectors are limited in several ways by our understanding of heavy-quark fragmentation.

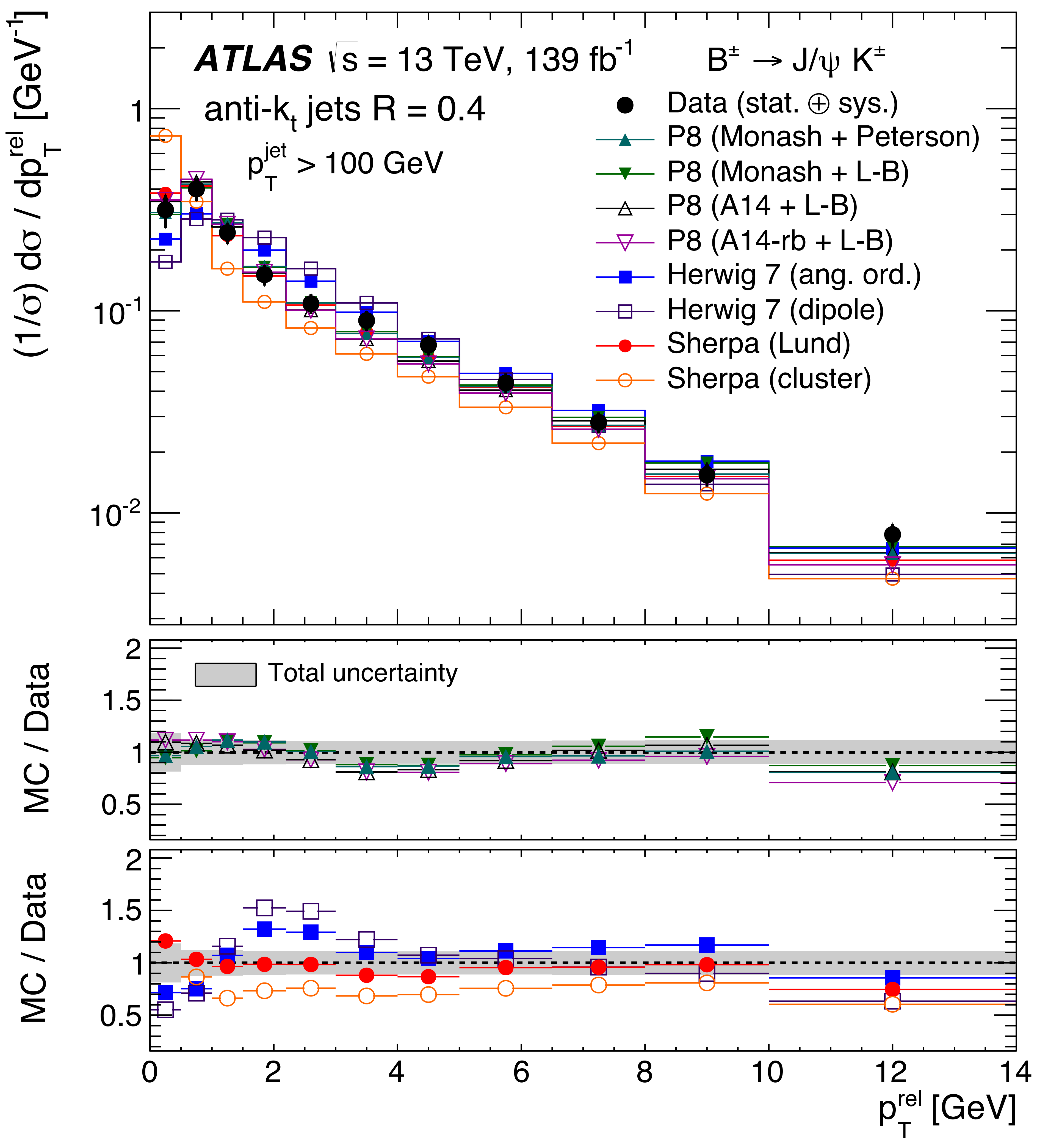
experimental methods at the LHC have been recently developed that can substantially improve this for the future.

there's still enormous room for these to grow in the coming years.
→ cannot take this for granted!

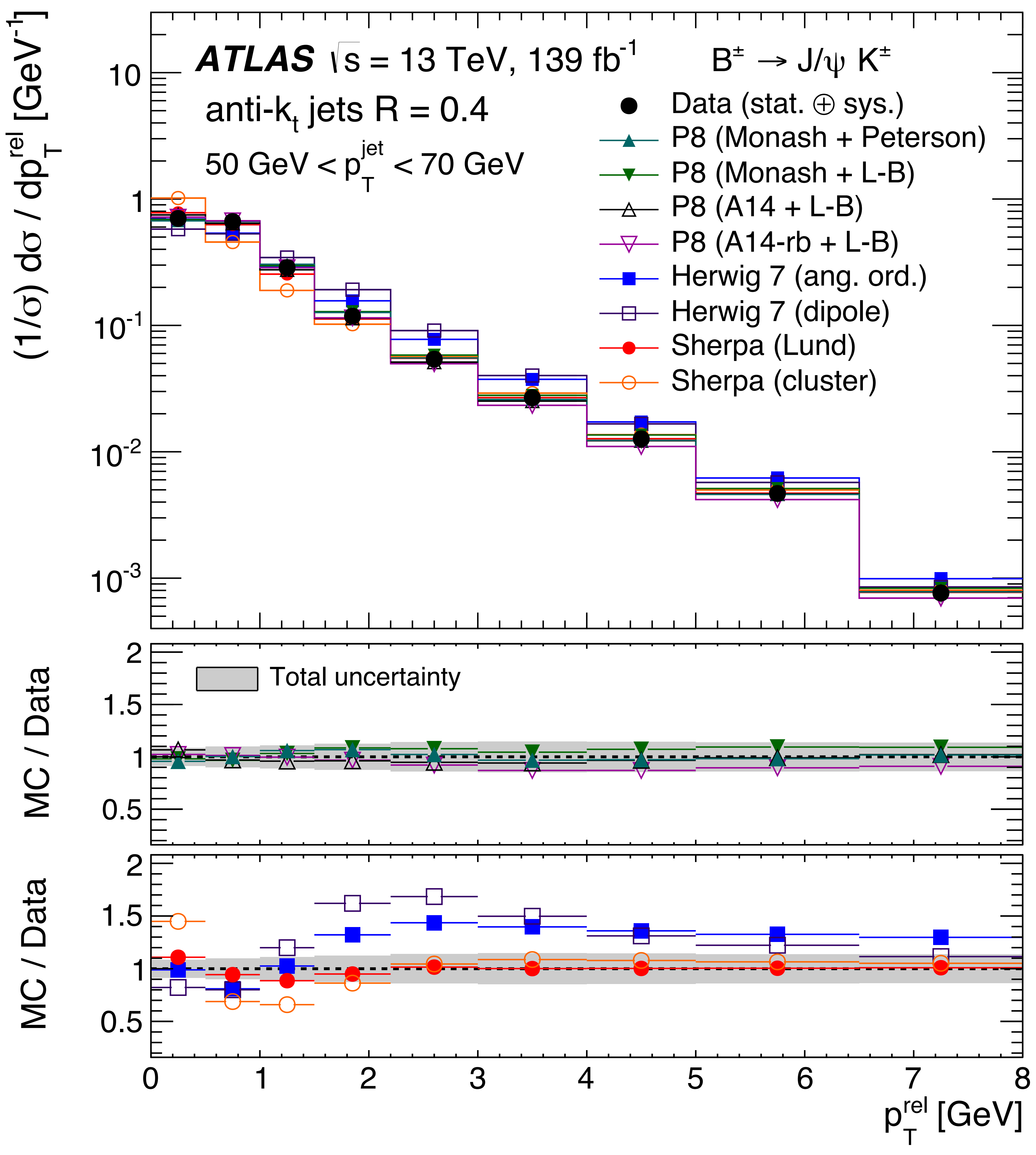
very happy for more interaction/ discussion with theory community.



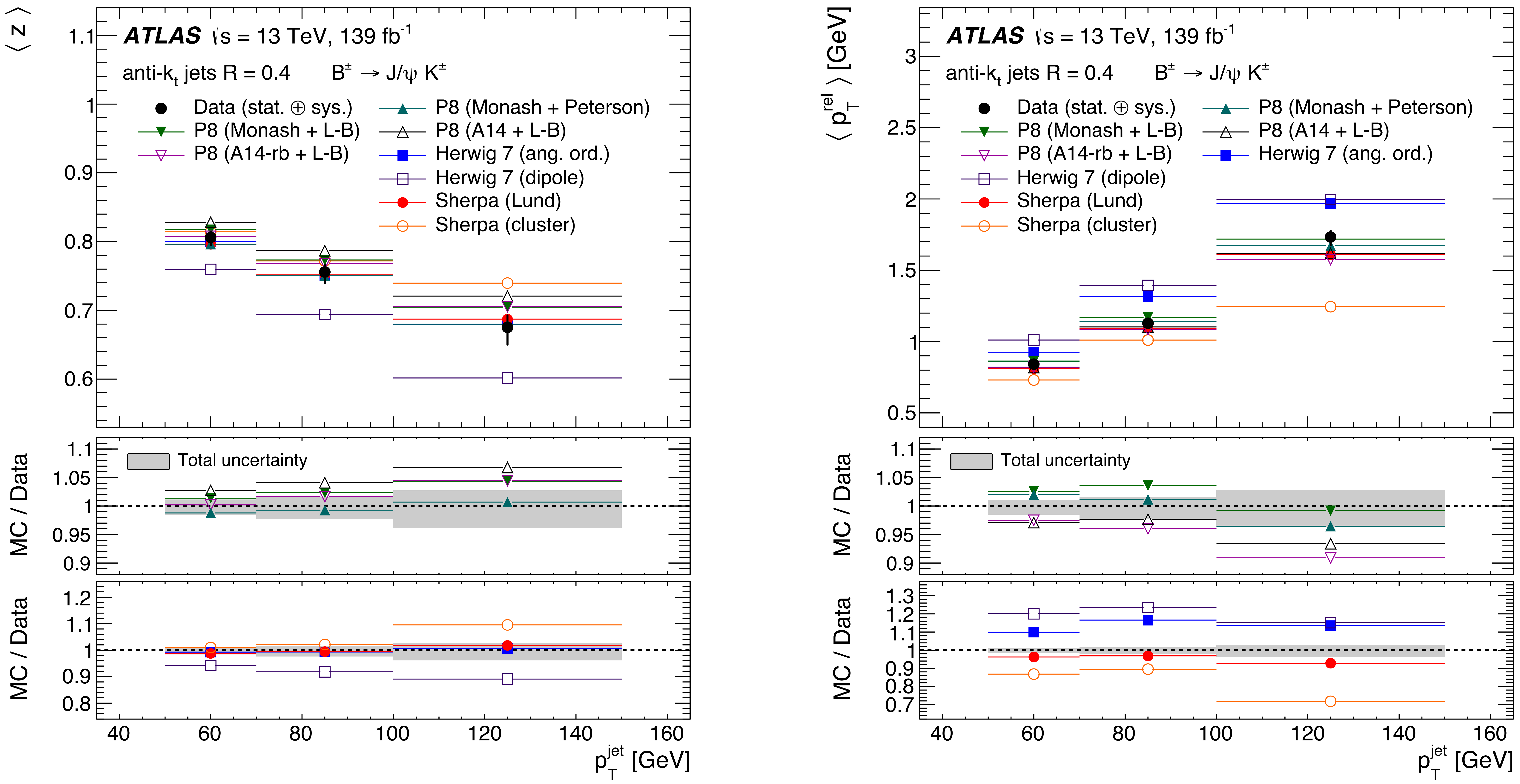
backup



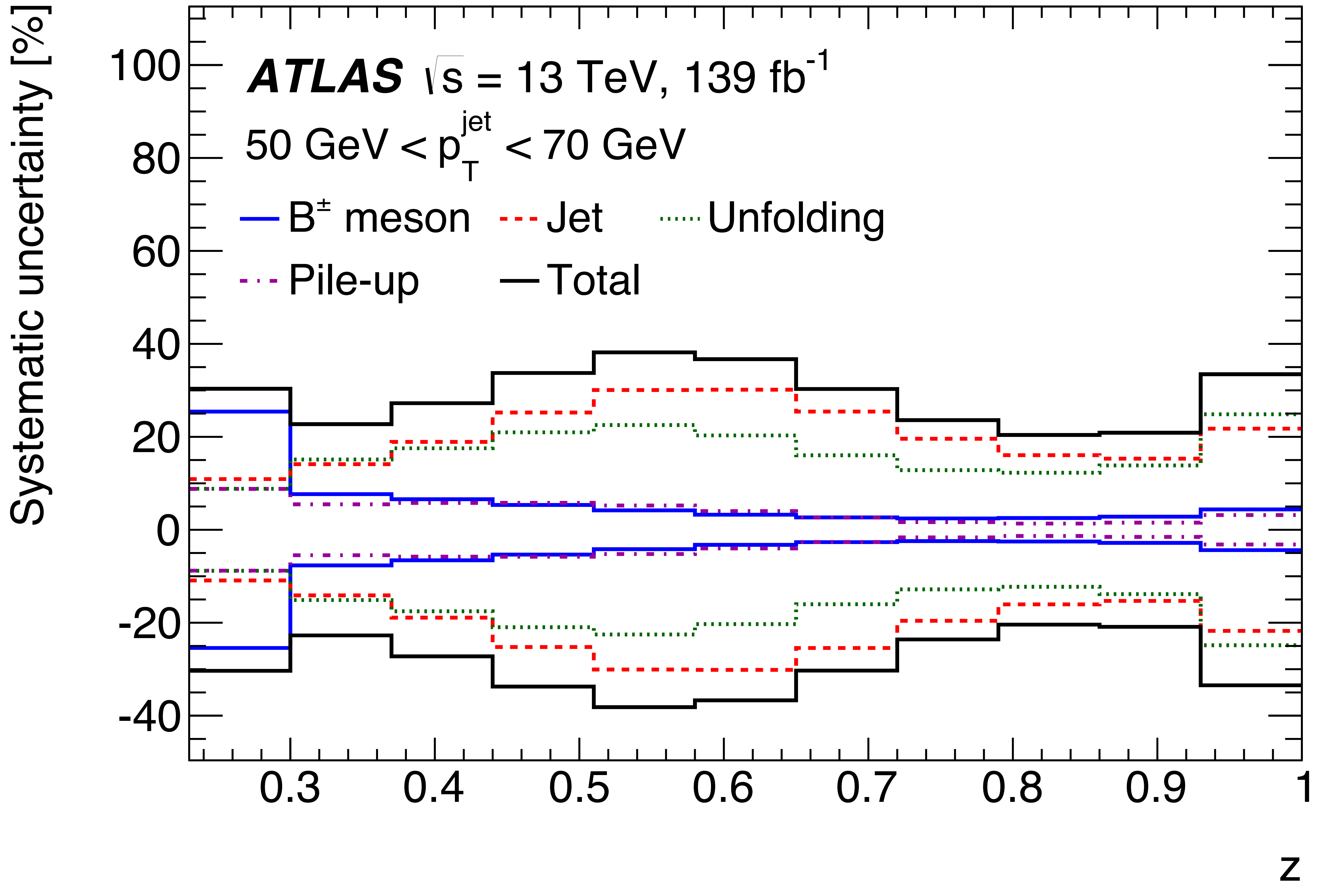
ATLAS √s = 13 TeV, 139 fb^-1
B± → J/ψ K±
anti-k_t jets R = 0.4
p_T^jet > 100 GeV
Data (stat. ⊕ sys.)
P8 (Monash + Peterson)
P8 (Monash + L-B)
P8 (A14 + L-B)
P8 (A14-rb + L-B)
Herwig 7 (ang. ord.)
Herwig 7 (dipole)
Sherpa (Lund)
Sherpa (cluster)
(1/σ) dσ / dp_T^rel [GeV^-1]
Total uncertainty
MC / Data
p_T^rel [GeV]

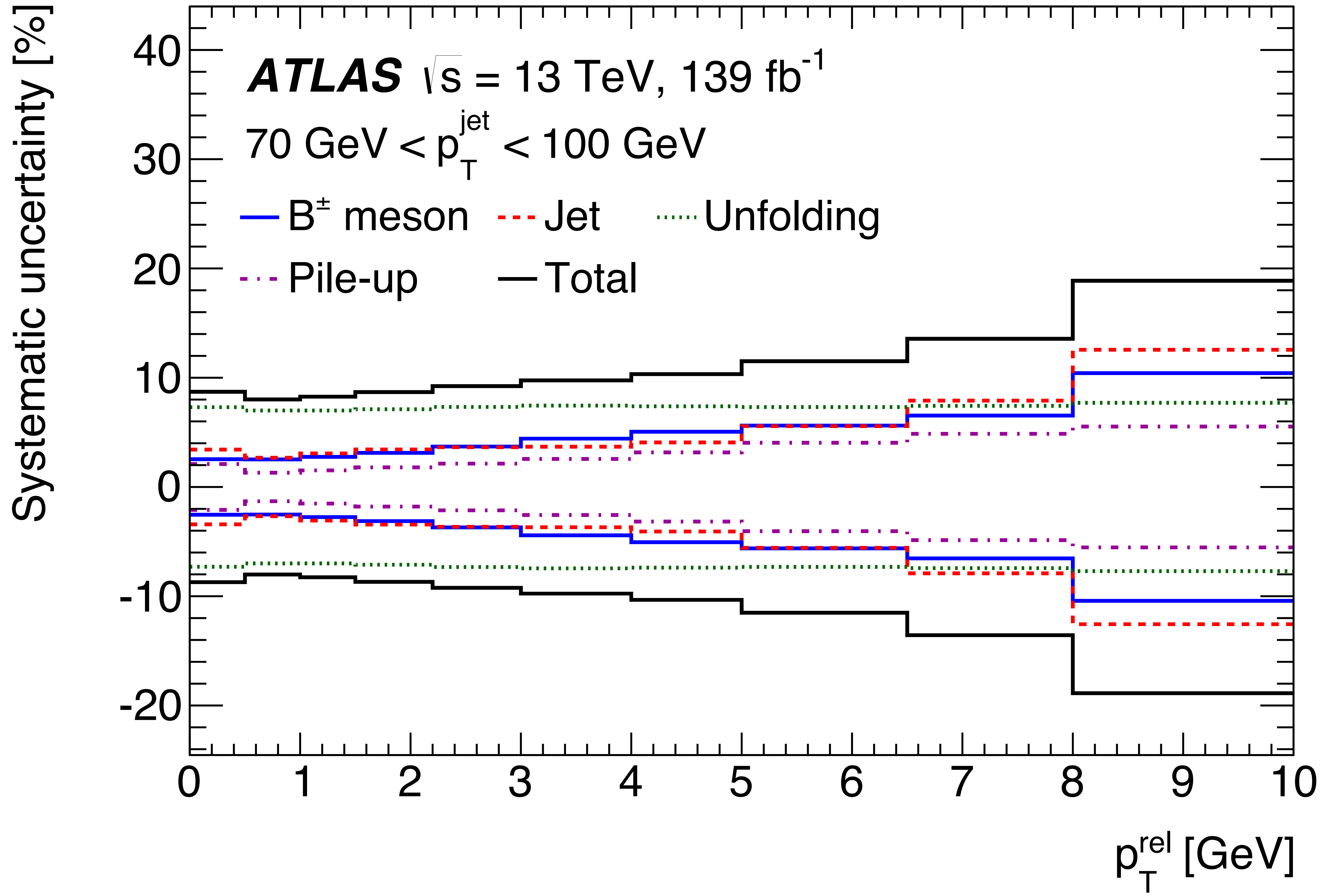


ATLAS √s = 13 TeV, 139 fb^-1
B± → J/ψ K±
anti-k_t jets R = 0.4
50 GeV < p_T^jet < 70 GeV
Data (stat. ⊕ sys.)
P8 (Monash + Peterson)
P8 (Monash + L-B)
P8 (A14 + L-B)
P8 (A14-rb + L-B)
Herwig 7 (ang. ord.)
Herwig 7 (dipole)
Sherpa (Lund)
Sherpa (cluster)
(1/σ) dσ / dp_T^rel [GeV^-1]
Total uncertainty
MC / Data
p_T^rel [GeV]

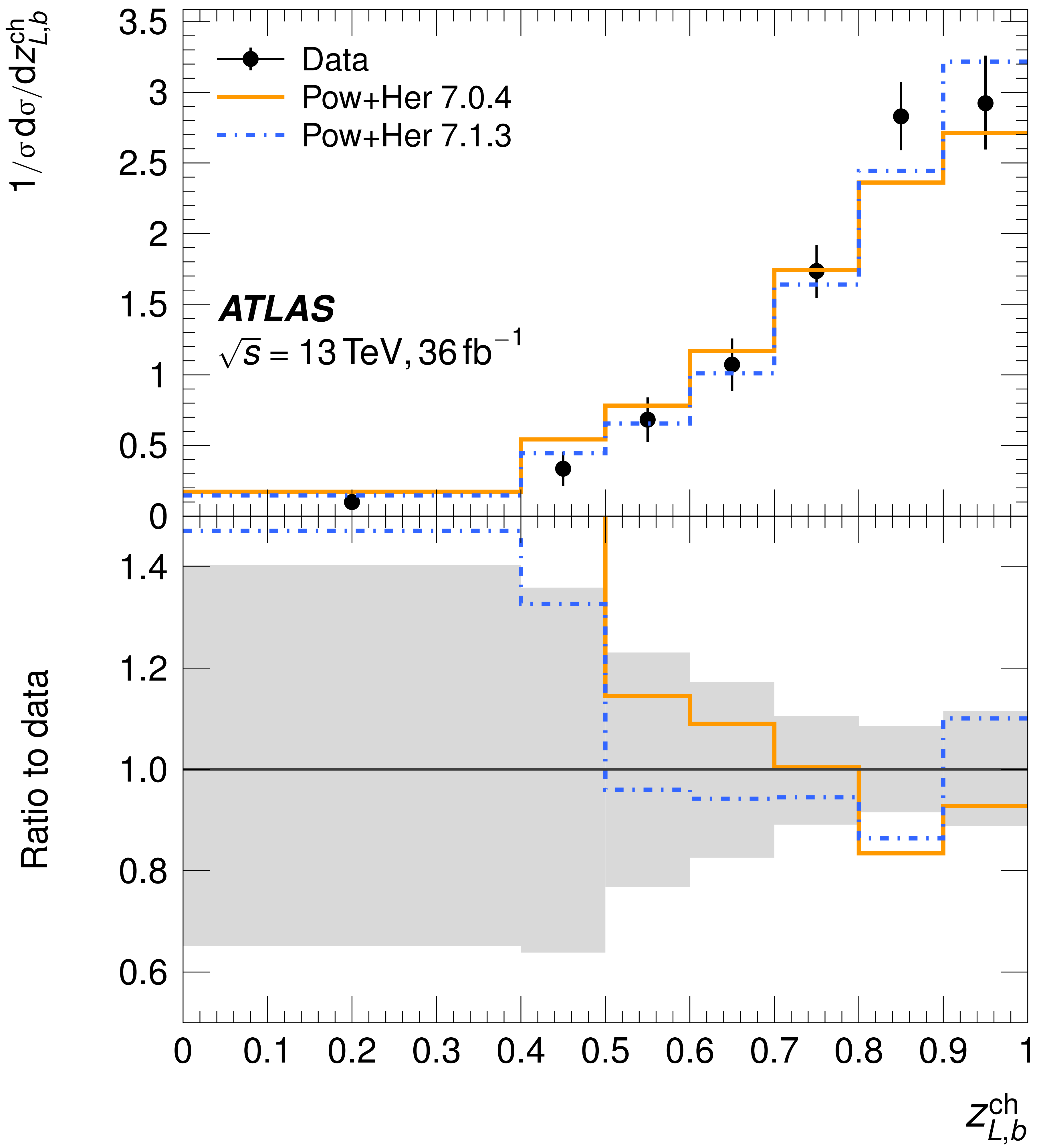


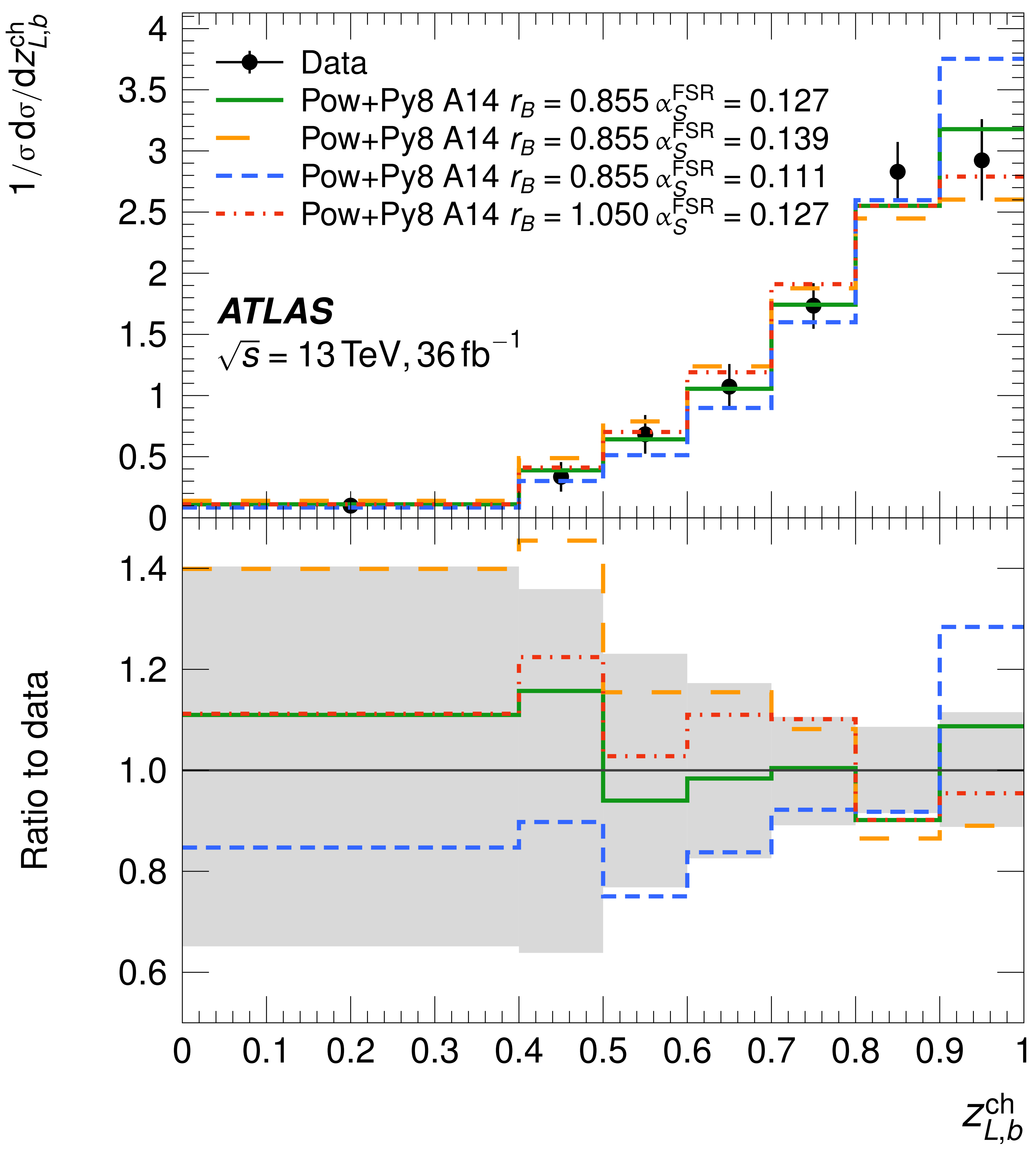
ATLAS √s = 13 TeV, 139 fb-1
anti-kt jets R = 0.4
B± → J/ψ K±
Data (stat. ⊕ sys.)
P8 (Monash + L-B)
P8 (A14-rb + L-B)
P8 (Monash + Peterson)
P8 (A14 + L-B)
Herwig 7 (ang. ord.)
Herwig 7 (dipole)
Sherpa (Lund)
Sherpa (cluster)
Total uncertainty
⟨ z ⟩
⟨ p_T^rel ⟩ [GeV]
MC / Data
p_T^jet [GeV]

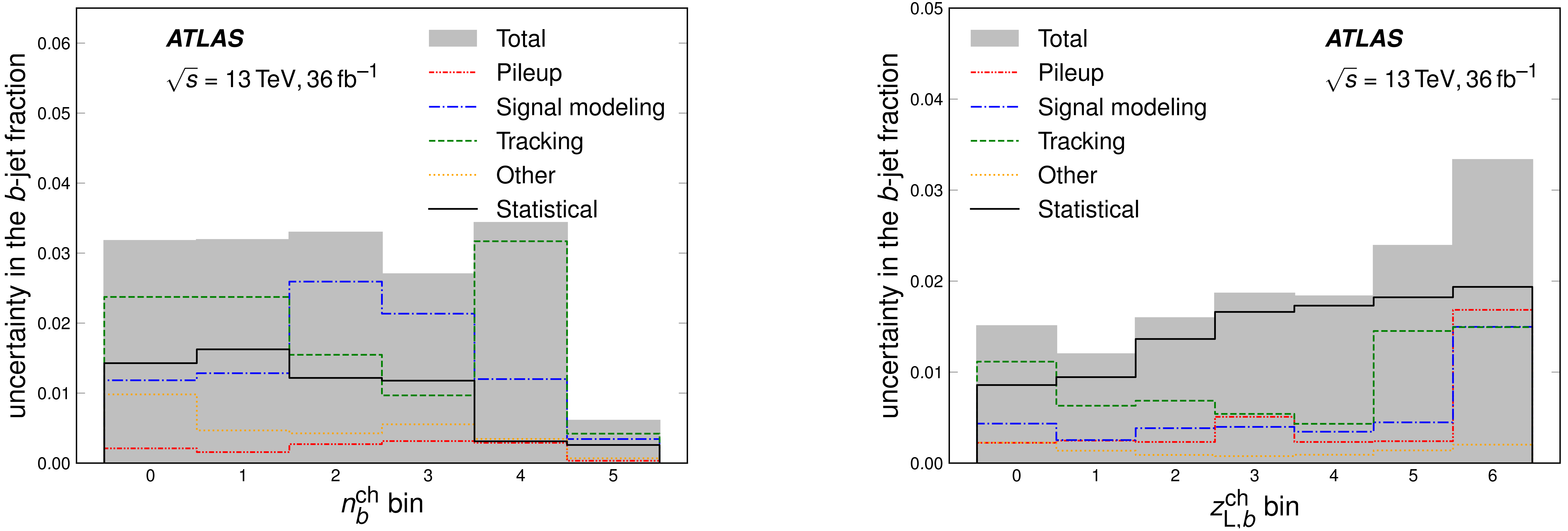

ATLAS √s = 13 TeV, 139 fb⁻¹
50 GeV < p_T^{jet} < 70 GeV
B± meson
Jet
Unfolding
Pile-up
Total
Systematic uncertainty [%]
z


ATLAS √s = 13 TeV, 139 fb⁻¹
70 GeV < p_T^{jet} < 100 GeV
B± meson
Jet
Unfolding
Pile-up
Total
Systematic uncertainty [%]
p_T^{rel} [GeV]

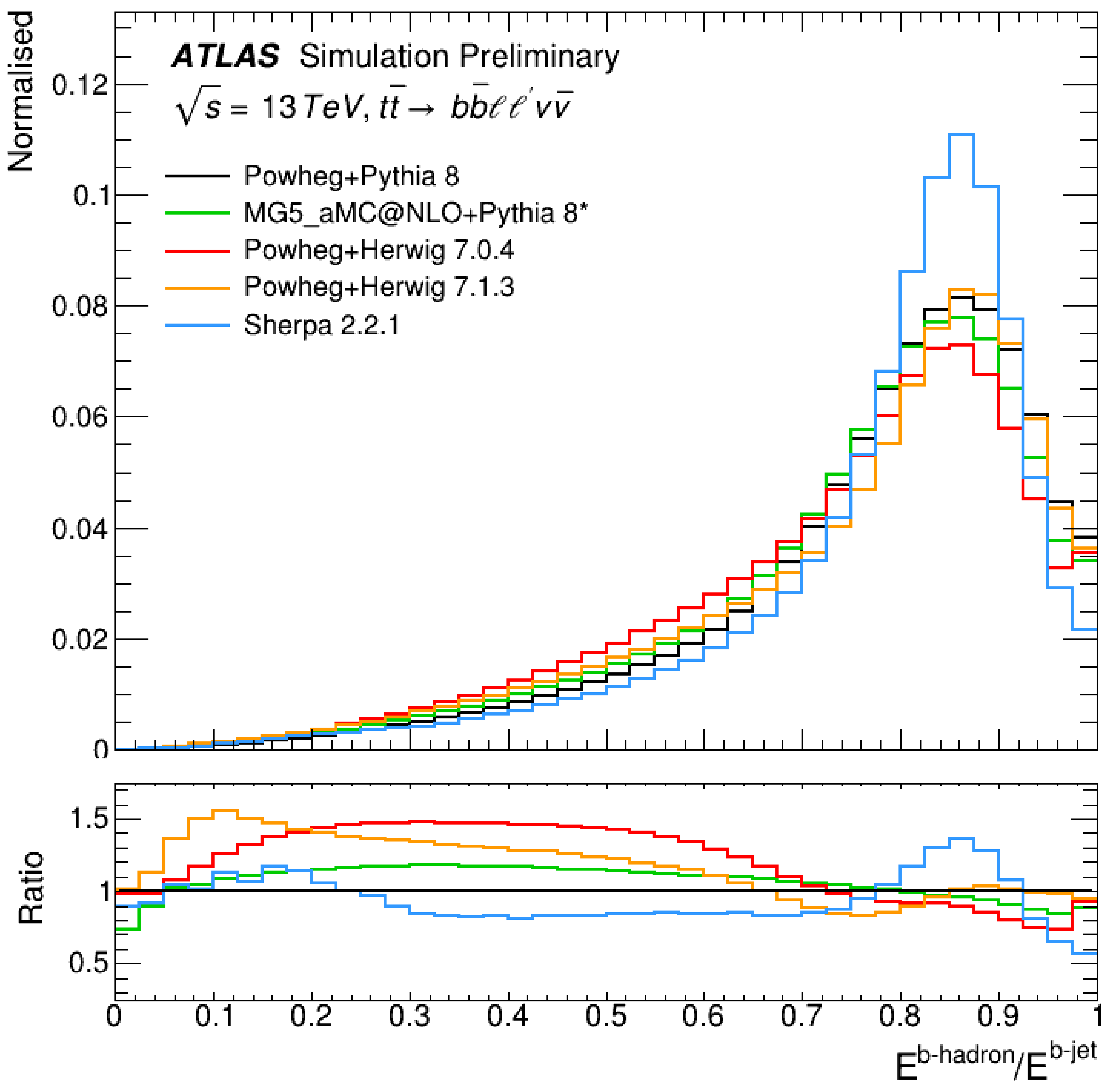
Data
Pow+Py8 A14 $r_B = 0.855$ $\alpha_S^{FSR} = 0.127$
Pow+Her 7.1.3
Sherpa 2.2.10
ATLAS
$\sqrt{s} = 13$ TeV, 36 fb^{-1}
Ratio to data
$z^{ch}_{L,b}$
$1/\sigma\, d\sigma/d\rho$
ρ
$1/\sigma\, d\sigma/dn^{ch}_b$
n^{ch}_b

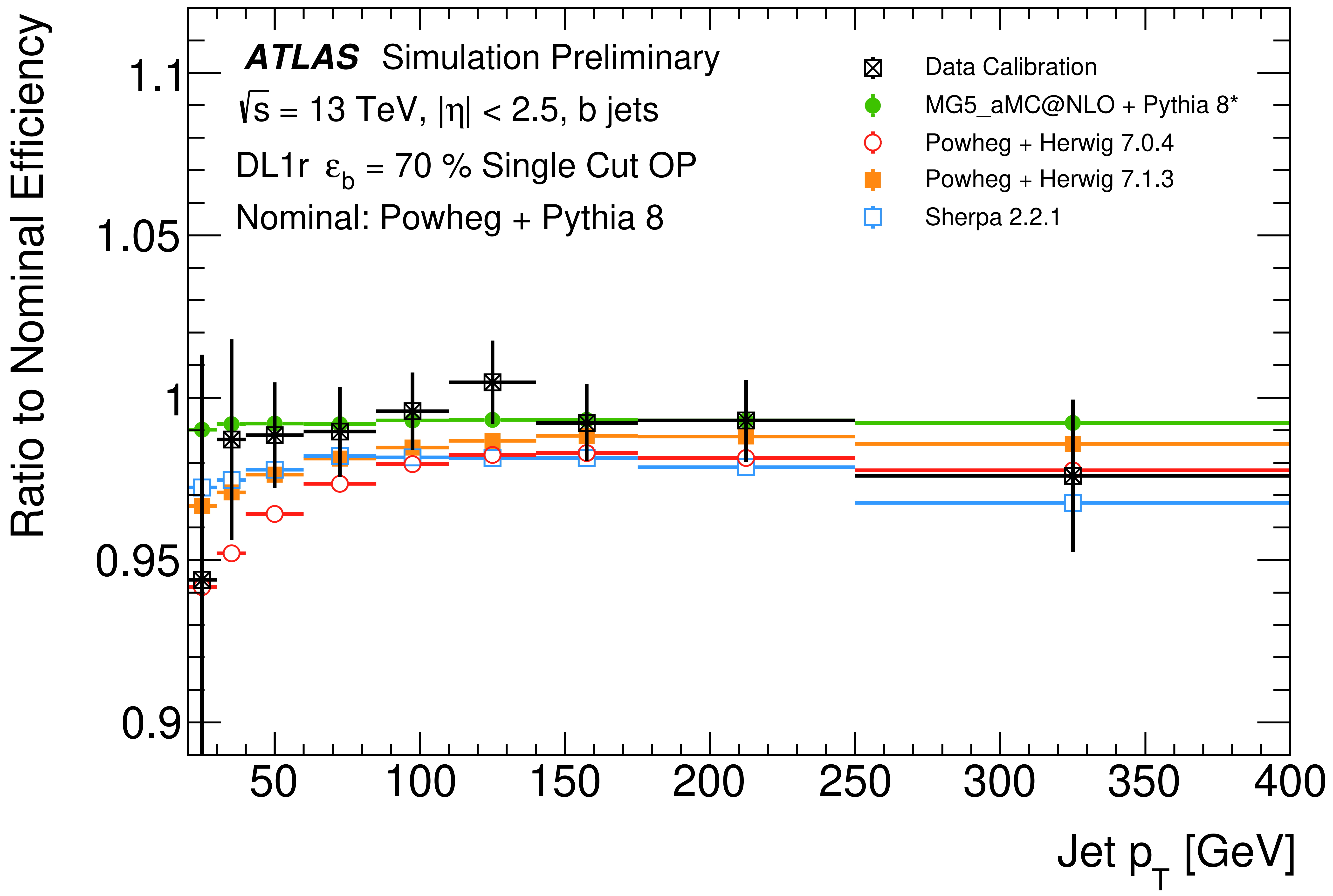

Data
Pow+Her 7.0.4
Pow+Her 7.1.3
ATLAS
$\sqrt{s}$ = 13 TeV, 36 fb^{-1}
$1/\sigma\, d\sigma/dz^{ch}_{L,b}$
Ratio to data
$z^{ch}_{L,b}$

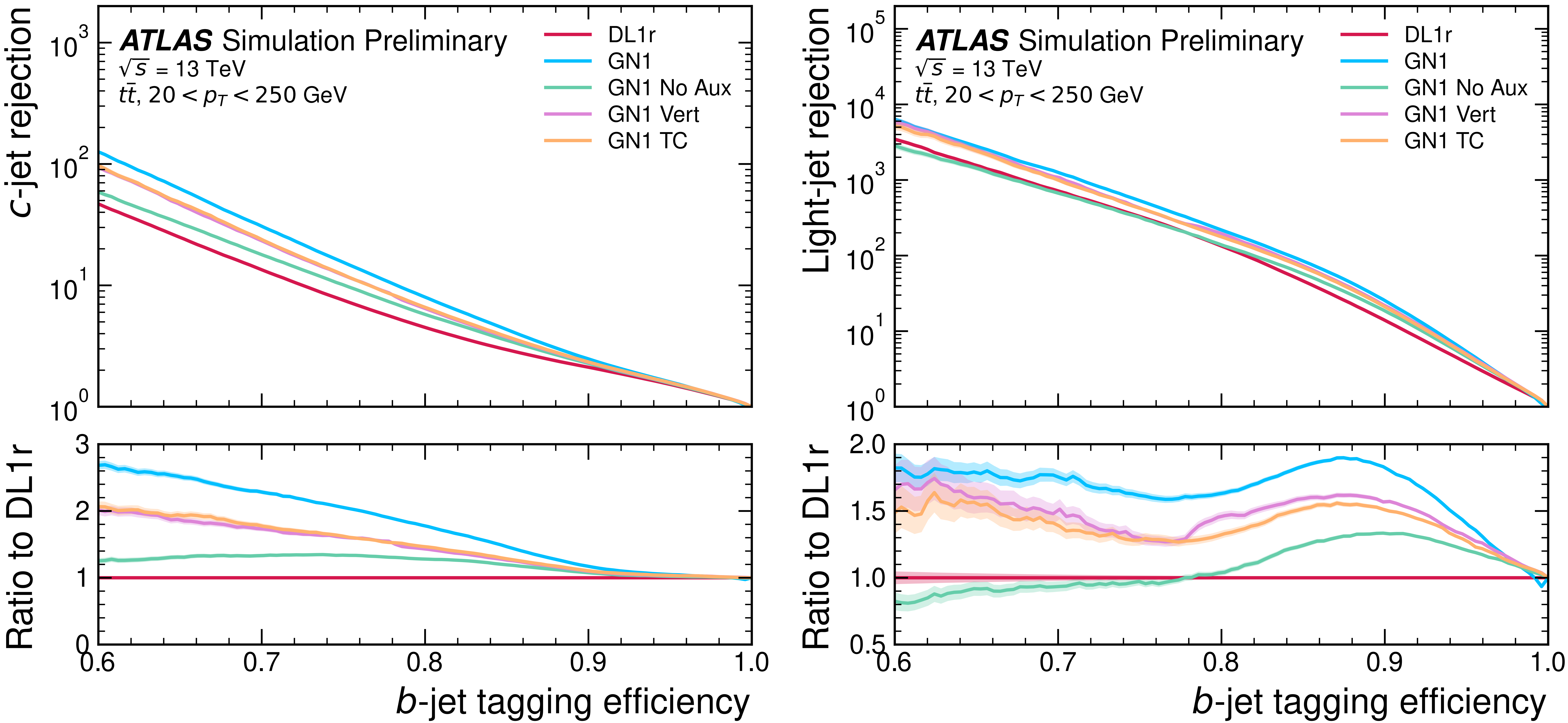

Data
Pow+Py8 A14 r_B = 0.855 α_S^{FSR} = 0.127
Pow+Py8 A14 r_B = 0.855 α_S^{FSR} = 0.139
Pow+Py8 A14 r_B = 0.855 α_S^{FSR} = 0.111
Pow+Py8 A14 r_B = 1.050 α_S^{FSR} = 0.127
ATLAS
$\sqrt{s}$ = 13 TeV, 36 fb^{-1}
$1/\sigma\, d\sigma/dz^{ch}_{L,b}$
Ratio to data
$z^{ch}_{L,b}$



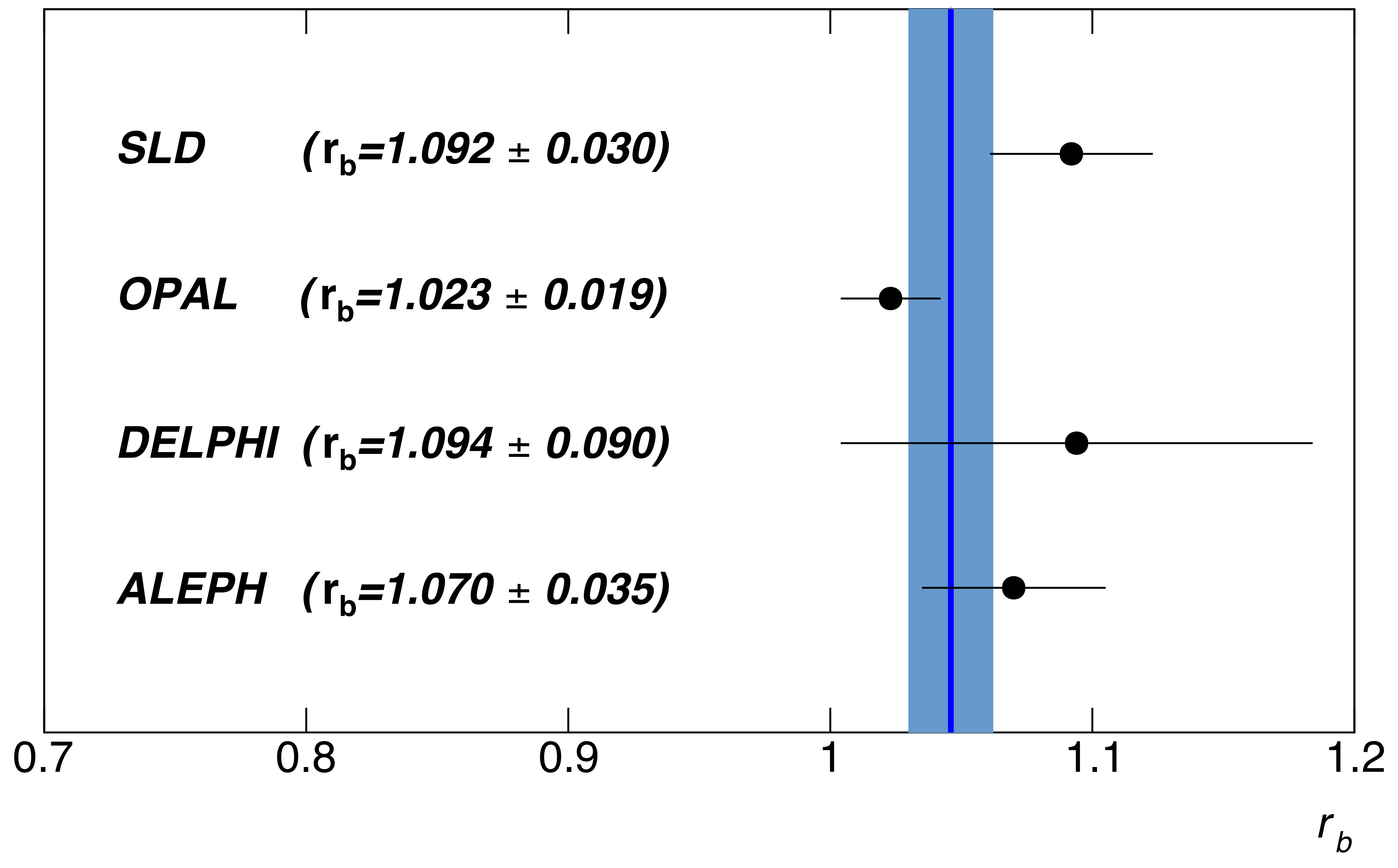
ATLAS
$\sqrt{s}$ = 13 TeV, 36 fb^{-1}
Total
Pileup
Signal modeling
Tracking
Other
Statistical
uncertainty in the b-jet fraction
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
5
n_b^{ch} bin
ATLAS
$\sqrt{s}$ = 13 TeV, 36 fb^{-1}
Total
Pileup
Signal modeling
Tracking
Other
Statistical
uncertainty in the b-jet fraction
0.00
0.01
0.02
0.03
0.04
0.05
0
1
2
3
4
5
6
$z_{L,b}^{ch}$ bin

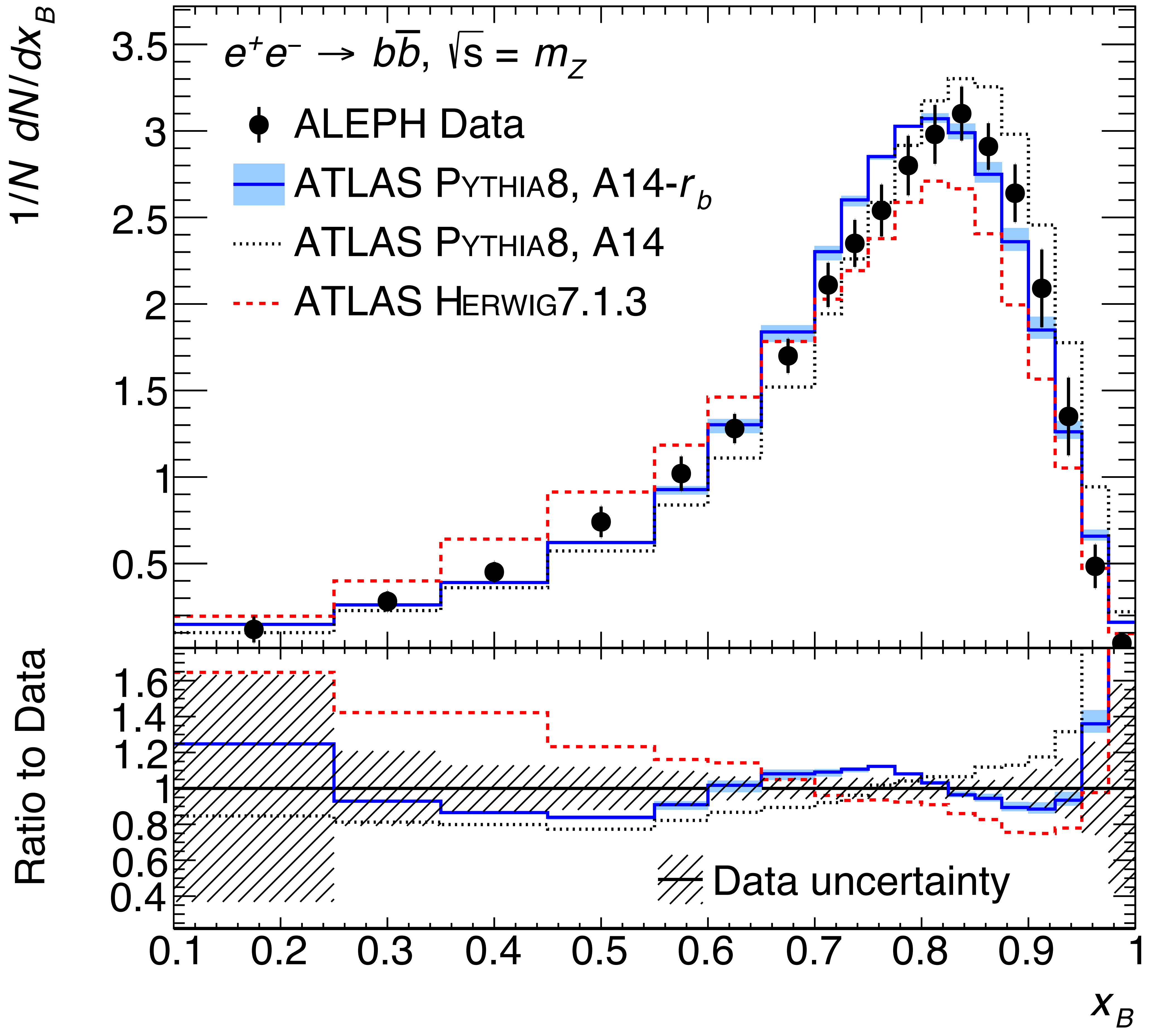

ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV, $t\bar{t} \rightarrow b\bar{b}\ell\ell'\nu\bar{\nu}$
Powheg+Pythia 8
MG5_aMC@NLO+Pythia 8*
Powheg+Herwig 7.0.4
Powheg+Herwig 7.1.3
Sherpa 2.2.1
Normalised
Ratio
$E^{b\text{-hadron}}/E^{b\text{-jet}}$

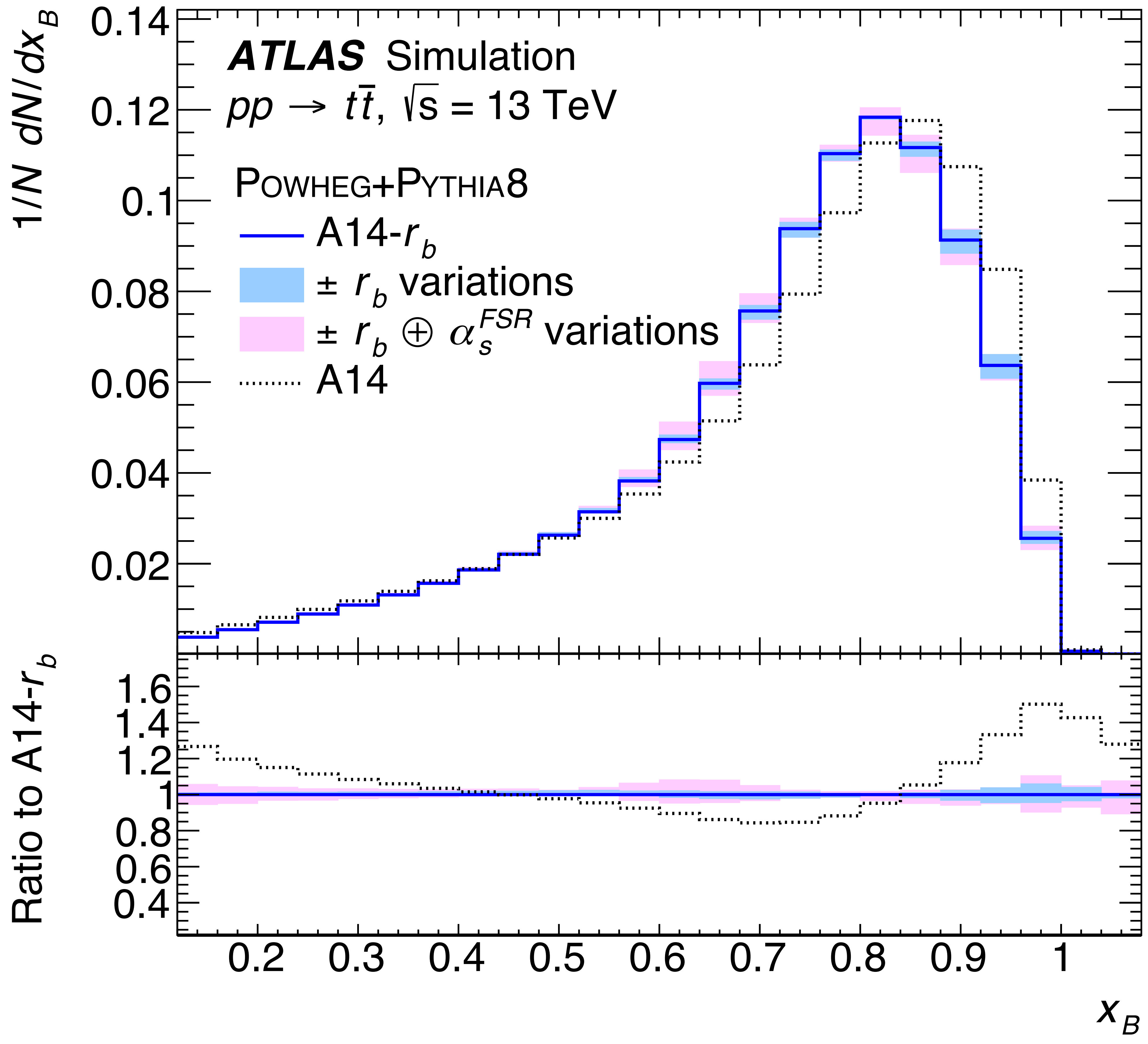

ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV, |η| < 2.5, b jets
DL1r ε_b = 70 % Single Cut OP
Nominal: Powheg + Pythia 8
Data Calibration
MG5_aMC@NLO + Pythia 8*
Powheg + Herwig 7.0.4
Powheg + Herwig 7.1.3
Sherpa 2.2.1
Ratio to Nominal Efficiency
Jet p_T [GeV]



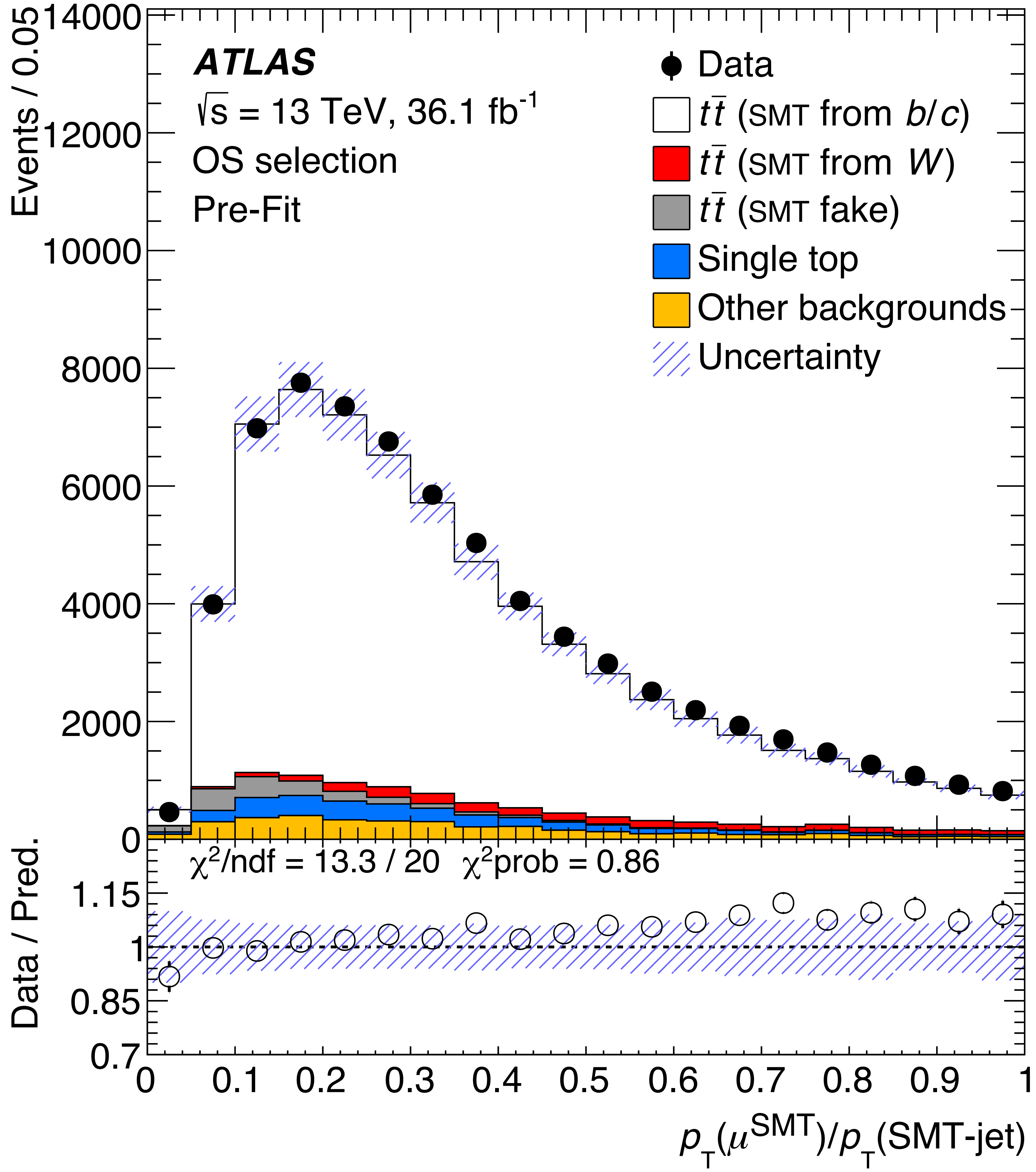
ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV
$t\bar{t}$, 20 < p_T < 250 GeV
DL1r
GN1
GN1 No Aux
GN1 Vert
GN1 TC
c-jet rejection
Light-jet rejection
Ratio to DL1r
b-jet tagging efficiency


SLD (r_b=1.092 ± 0.030)
OPAL (r_b=1.023 ± 0.019)
DELPHI (r_b=1.094 ± 0.090)
ALEPH (r_b=1.070 ± 0.035)
0.7
0.8
0.9
1
1.1
1.2
r_b

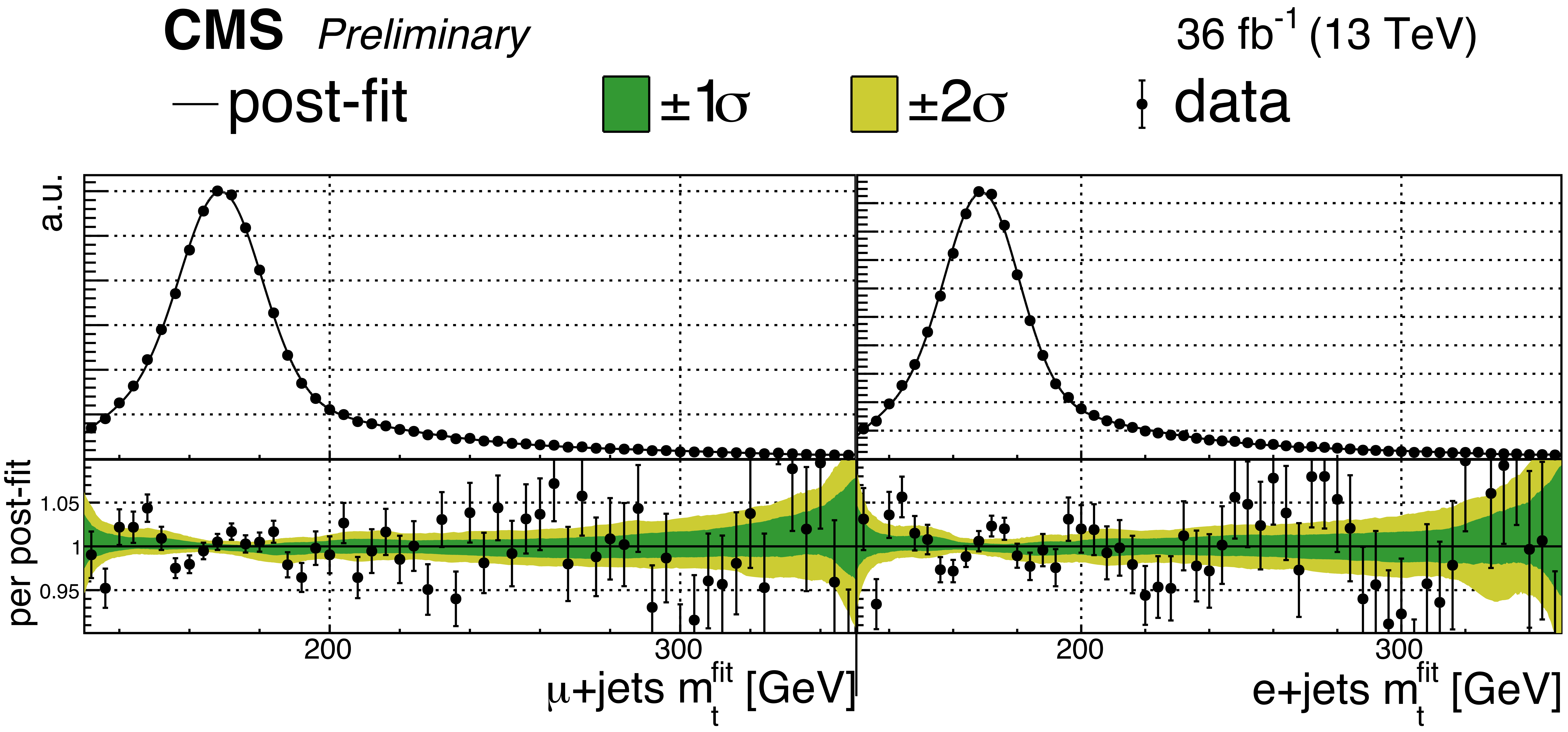

e+e− → bb̄, √s = m_Z
ALEPH Data
ATLAS PYTHIA8, A14-r_b
ATLAS PYTHIA8, A14
ATLAS HERWIG7.1.3
1/N dN/dx_B
Ratio to Data
Data uncertainty
x_B



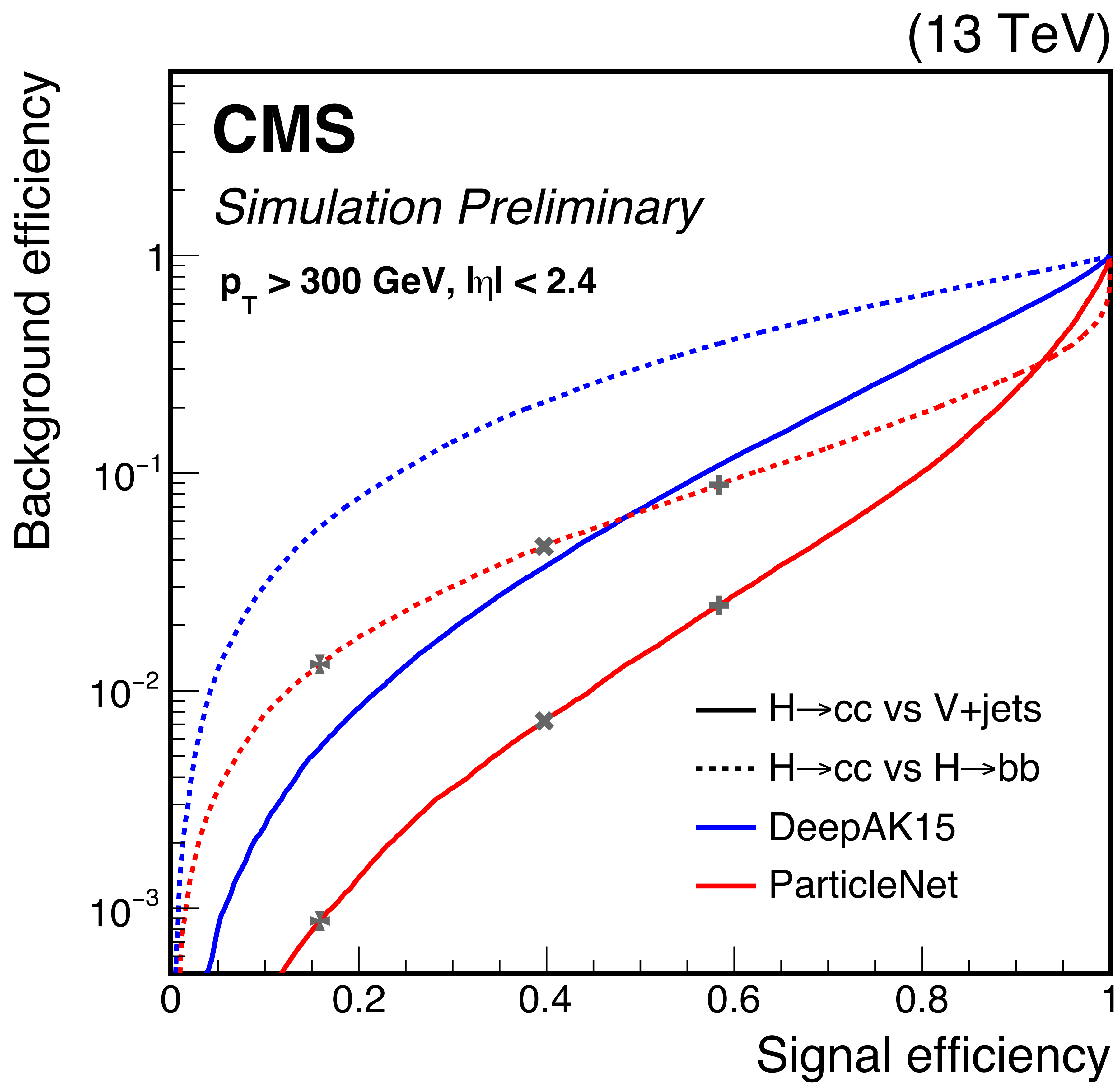
ATLAS Simulation
pp → $t\bar{t}$, √s = 13 TeV
POWHEG+PYTHIA8
A14-r_b
± r_b variations
± r_b ⊕ α_s^{FSR} variations
A14
1/N dN/dx_B
Ratio to A14-r_b
x_B

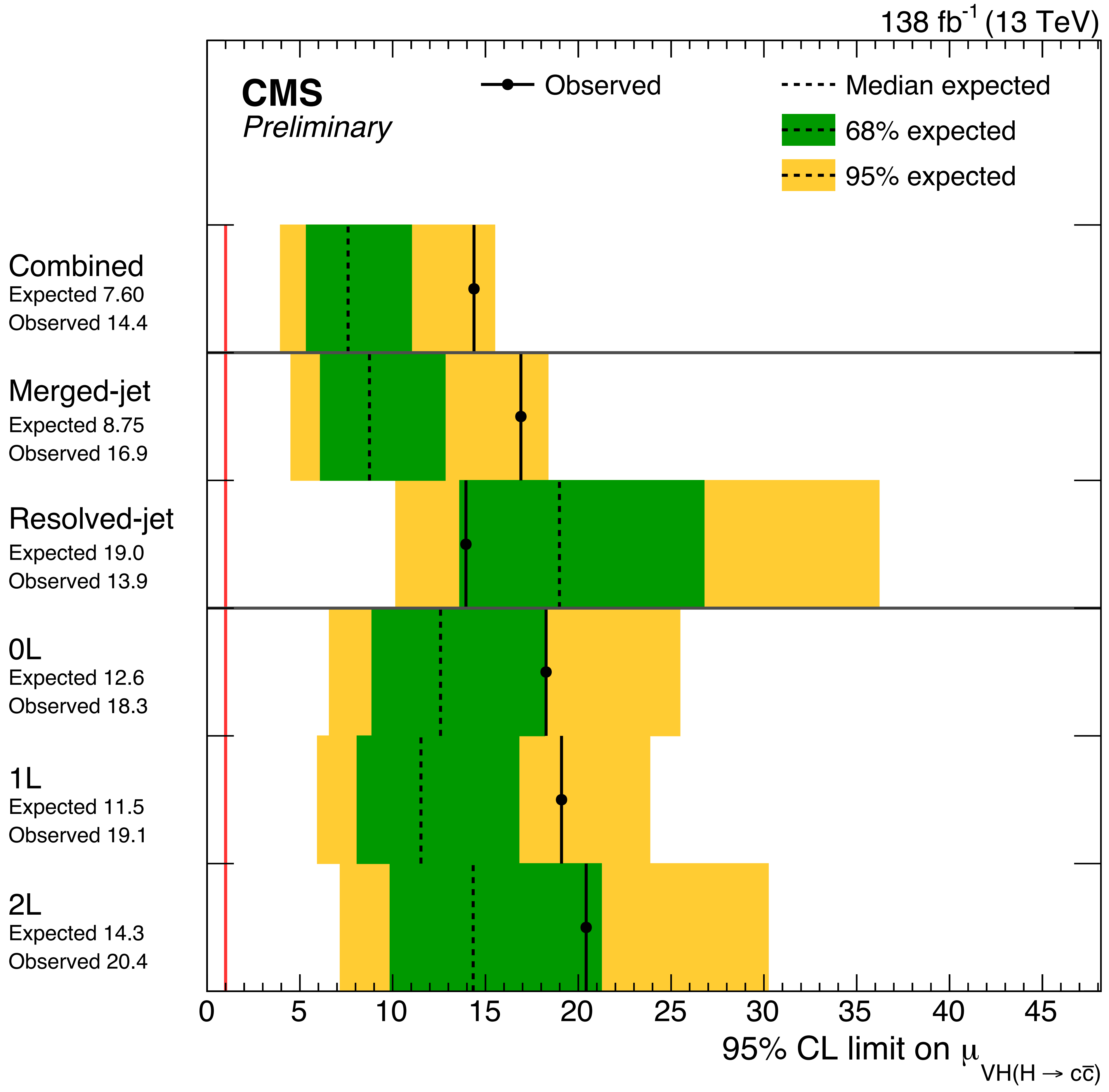


ATLAS
√s = 13 TeV, 36.1 fb^{-1}
OS selection
Pre-Fit
Data
$t\bar{t}$ (SMT from b/c)
$t\bar{t}$ (SMT from W)
$t\bar{t}$ (SMT fake)
Single top
Other backgrounds
Uncertainty
Events / 0.05
χ^2/ndf = 13.3 / 20 χ^2prob = 0.86
Data / Pred.
$p_T(\mu^{SMT})/p_T$(SMT-jet)



CMS Preliminary
36 fb-1 (13 TeV)
post-fit
±1σ
±2σ
data
a.u.
per post-fit
1.05
1
0.95
200
300
μ+jets m_t^fit [GeV]
e+jets m_t^fit [GeV]

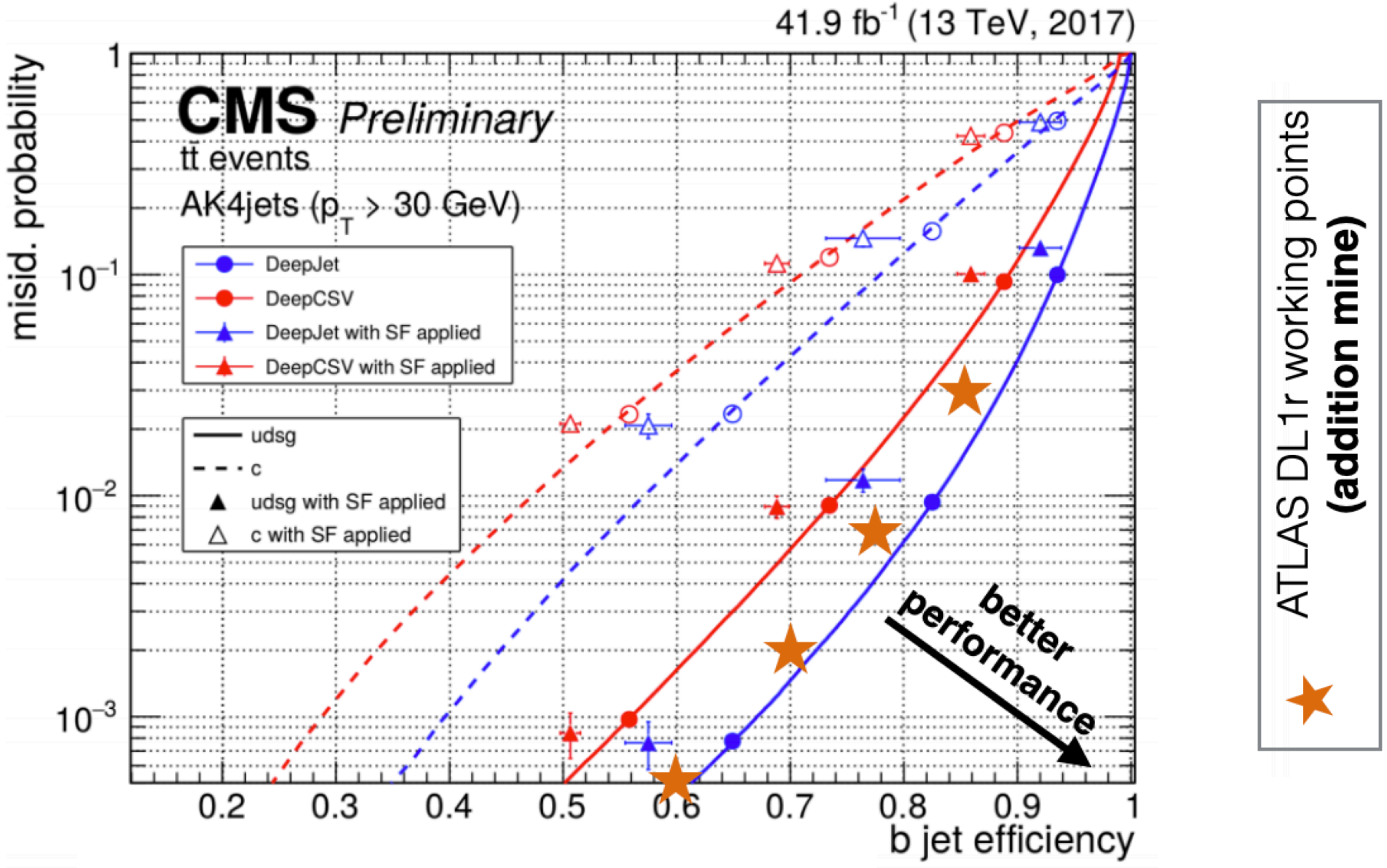

(13 TeV)
CMS
Simulation Preliminary
p_T > 300 GeV, |η| < 2.4
Background efficiency
1
10^−1
10^−2
10^−3
0
0.2
0.4
0.6
0.8
1
Signal efficiency
H→cc vs V+jets
H→cc vs H→bb
DeepAK15
ParticleNet

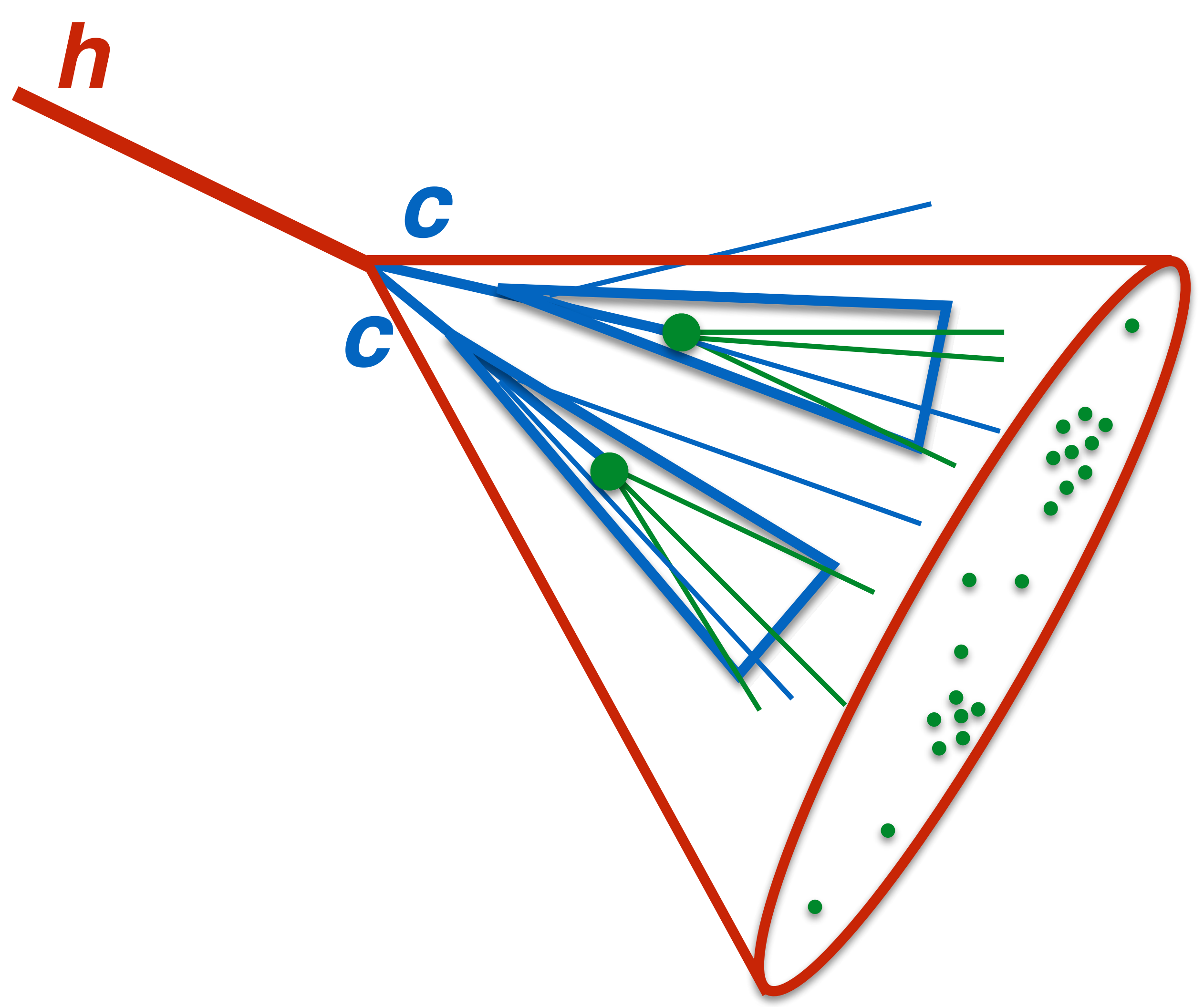

138 fb^-1 (13 TeV)
CMS
Preliminary
Observed
Median expected
68% expected
95% expected
Combined
Expected 7.60
Observed 14.4
Merged-jet
Expected 8.75
Observed 16.9
Resolved-jet
Expected 19.0
Observed 13.9
0L
Expected 12.6
Observed 18.3
1L
Expected 11.5
Observed 19.1
2L
Expected 14.3
Observed 20.4
0
5
10
15
20
25
30
35
40
45
95% CL limit on μ_VH(H → cc̄)

a cautionary tale
(apologies in advance to CMS)

in the machine-learning era, small mismodeling of b-jet internals has large consequences.

CMS were losing ~10% of signal b-jets in data c.f. simulation.

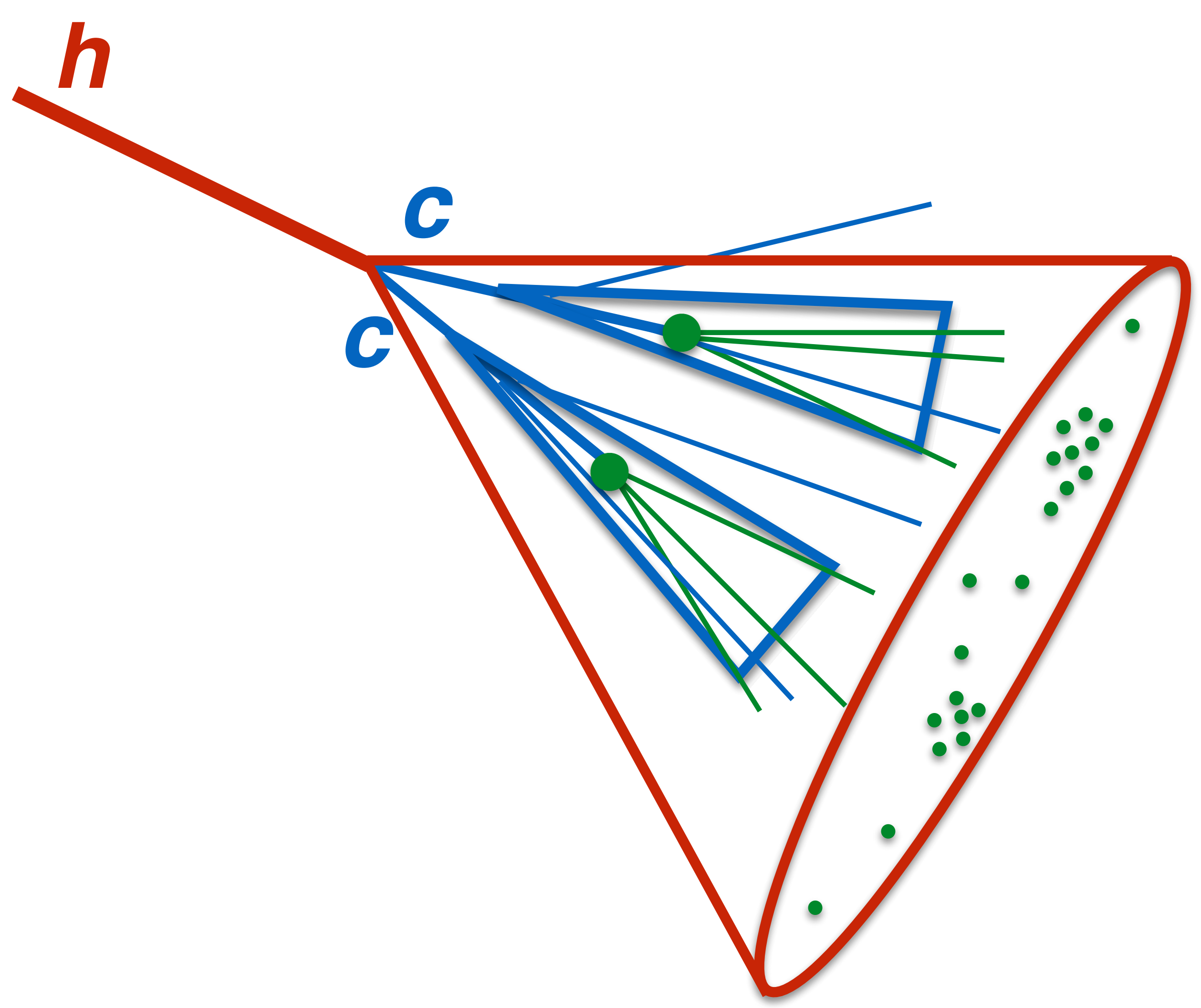


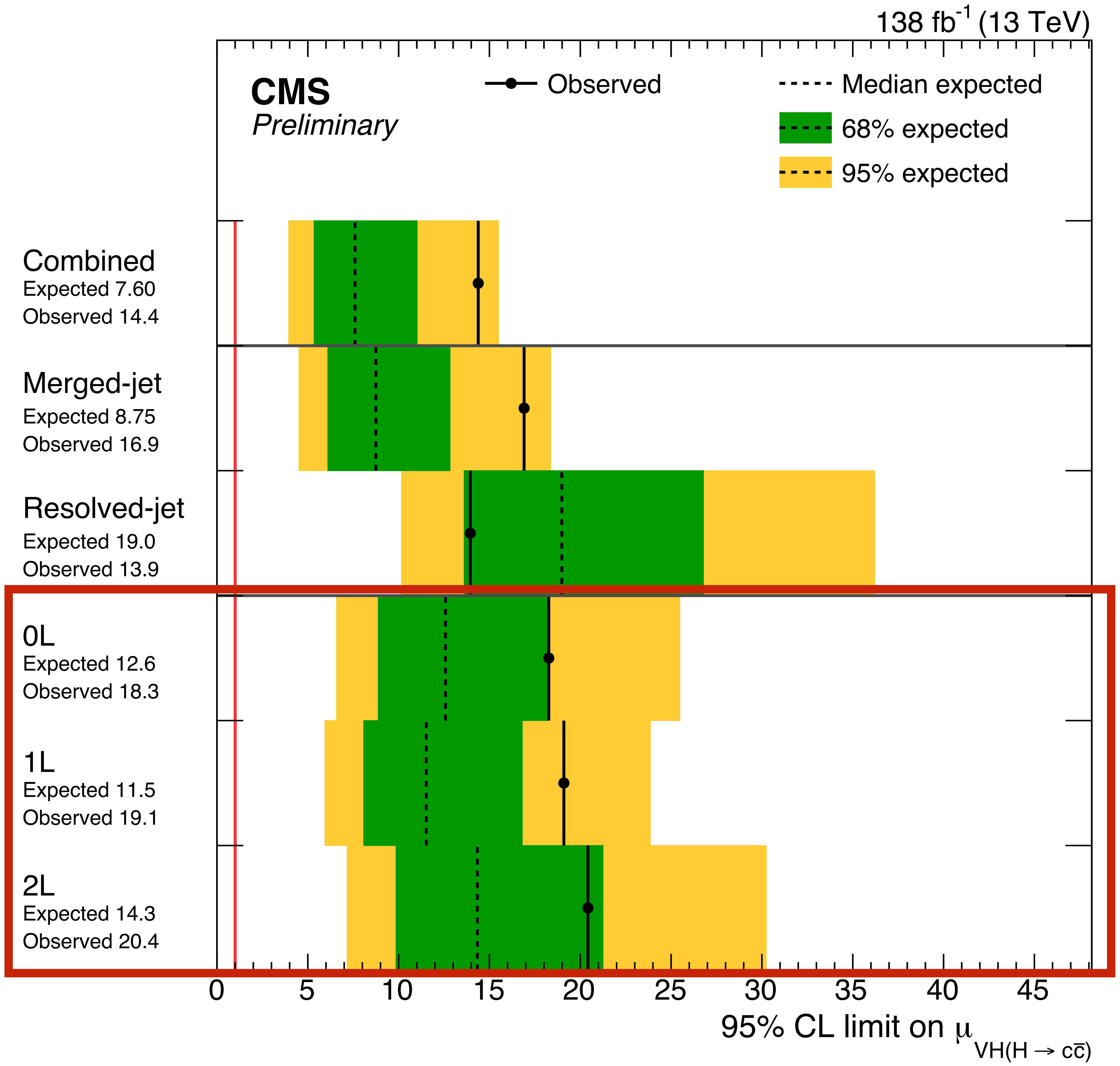


CMS have also shown very strong promise in tagging high-p_T $h \rightarrow cc$ decays, putting very fine-grained details into a NN discriminator.

However, we have very poor experimental constraints on charm hadron production.

This is a dangerous game…





CMS have also shown very strong promise in tagging high-p_T $h \to cc$ decays, putting very fine-grained details into a NN discriminator.

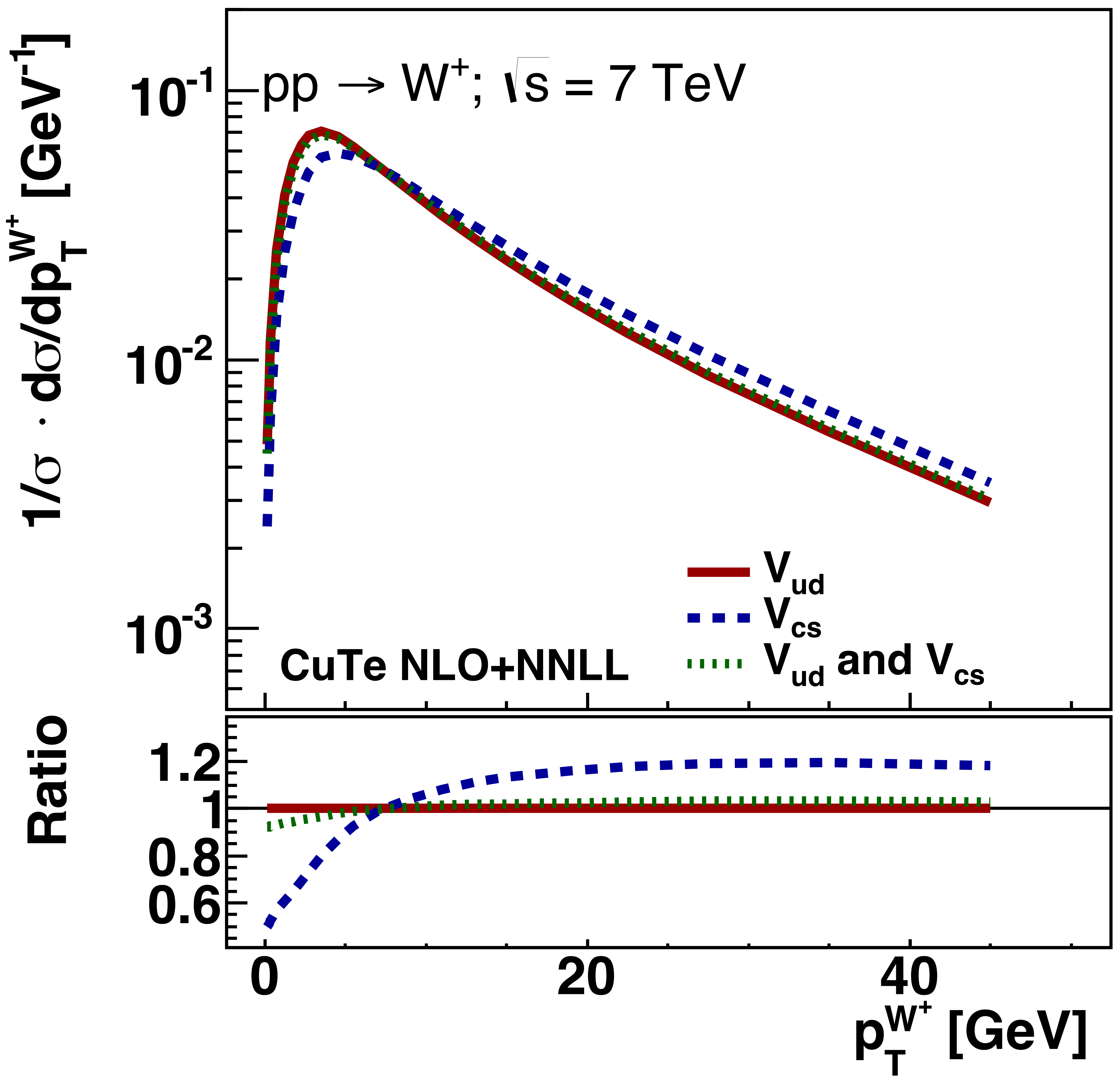
However, we have very poor experimental constraints on charm hadron production.

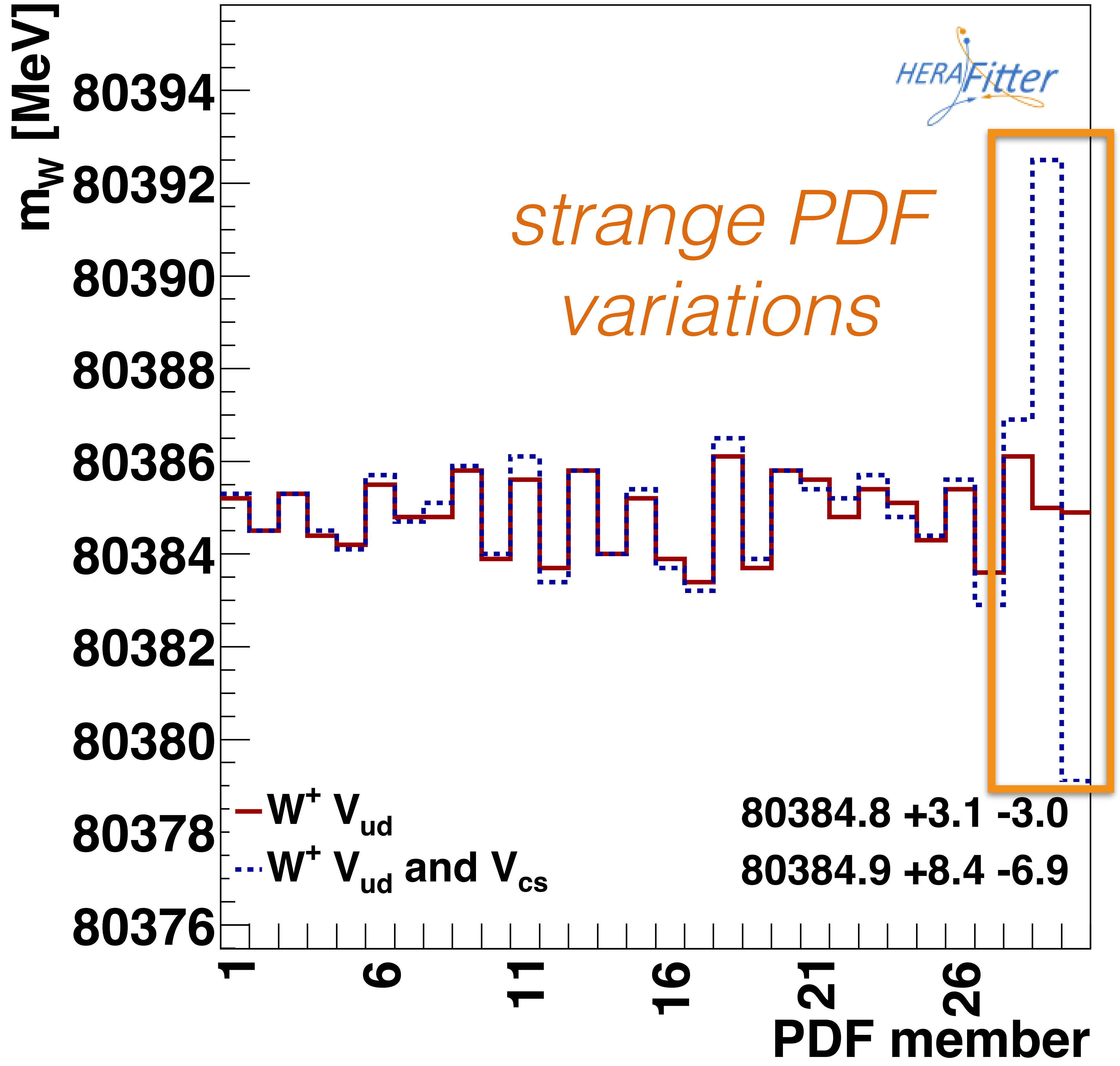
This is a dangerous game…

strong direct constraints on the strange-quark PDF come via measurements of $W + c$ production,

but c-quark fragmentation is limiting there, too.

big implications for W mass measurements at the LHC.

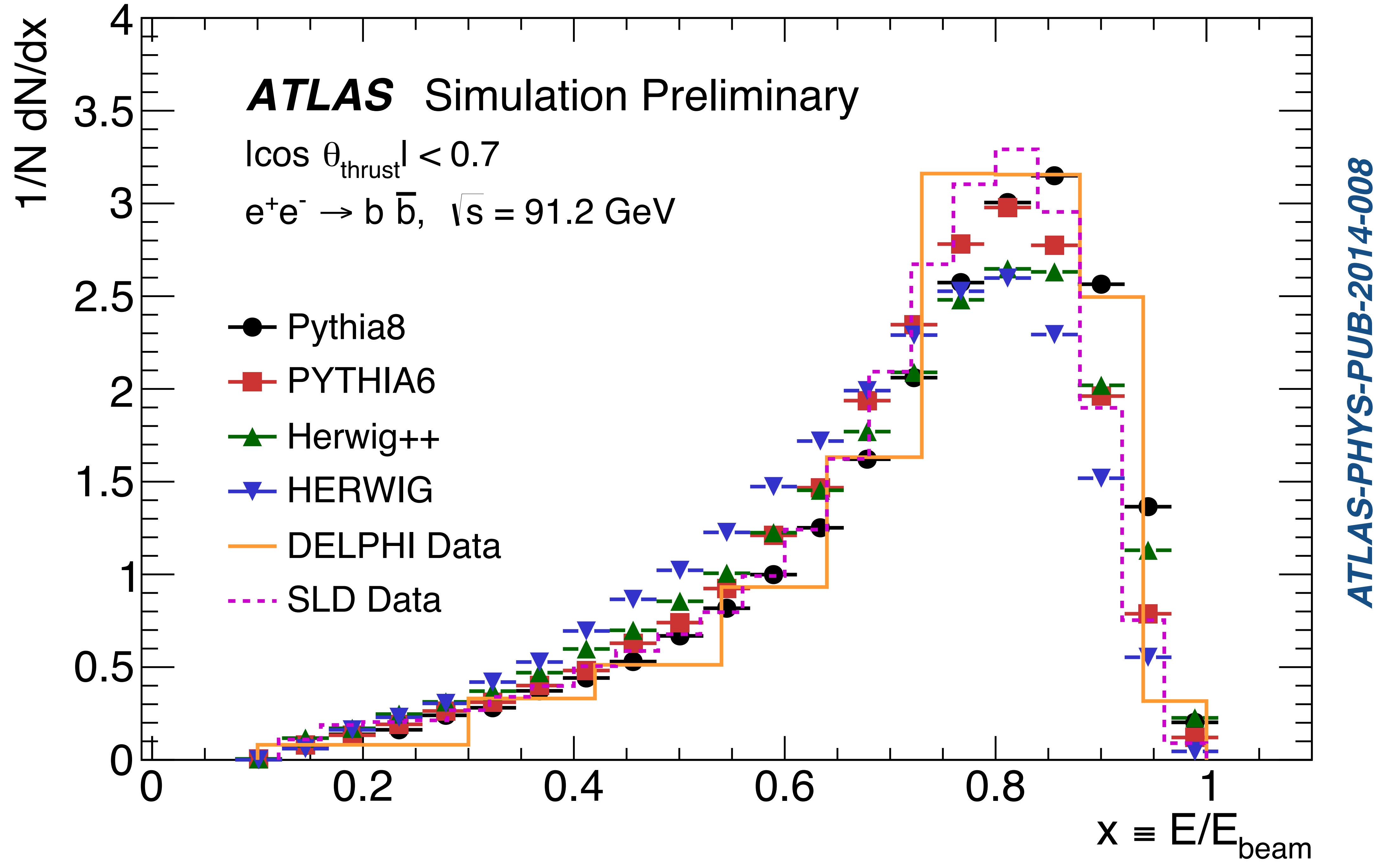




how?

- *b- and c-hadrons leave a striking experimental signature*
- *there is a unique correspondence to the originating heavy flavor quarks*
    - *ergo a precise probe of QCD*

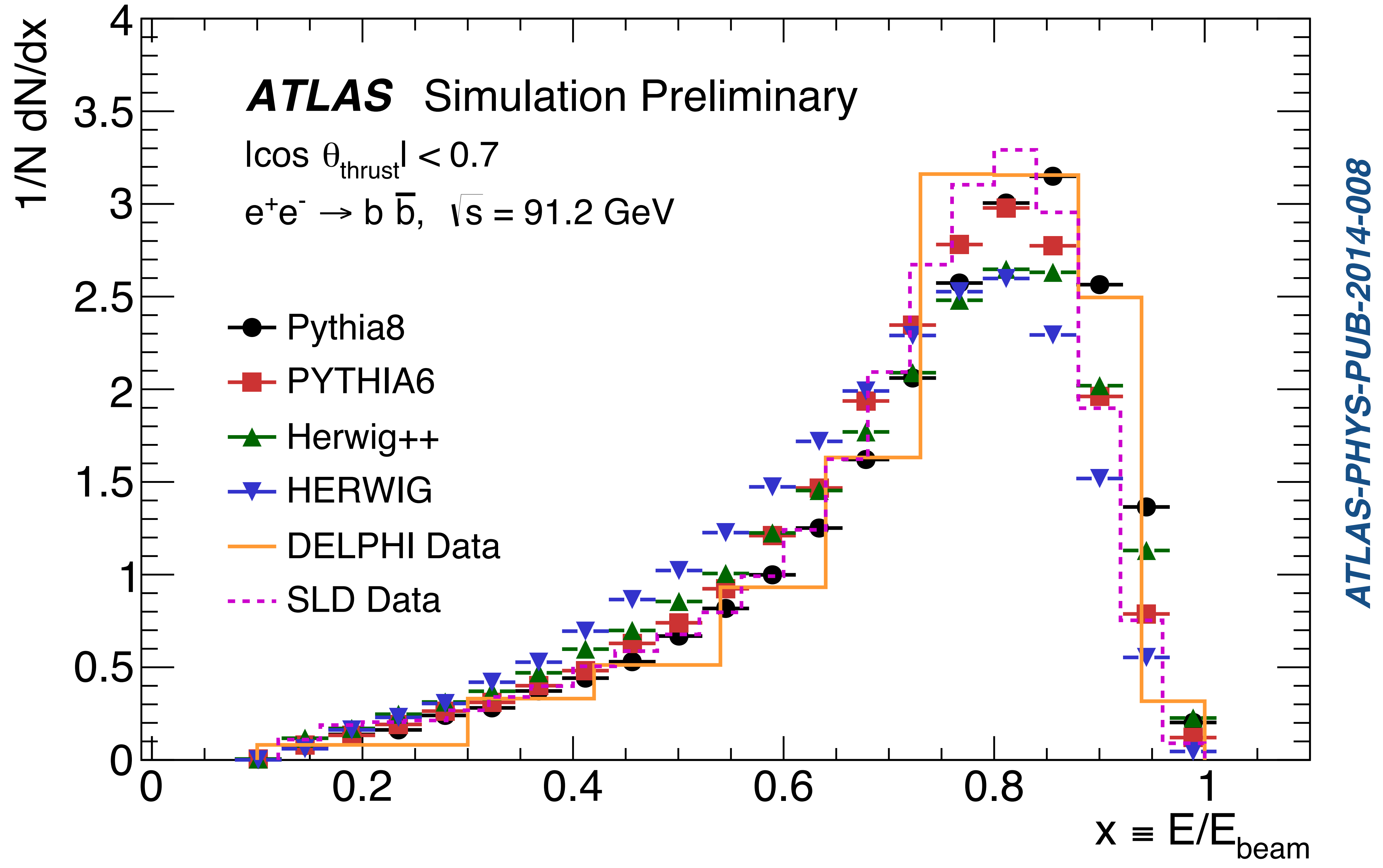
- b- and c-hadrons leave a striking experimental signature
- there is a unique correspondence to the originating heavy flavor quarks
  - ergo a precise probe of QCD
- b-fragmentation currently tuned to e^+e^- data from $Z \to bb$ decays.



tension between $e^+e^- \to bb$ measurements of b-fragmentation

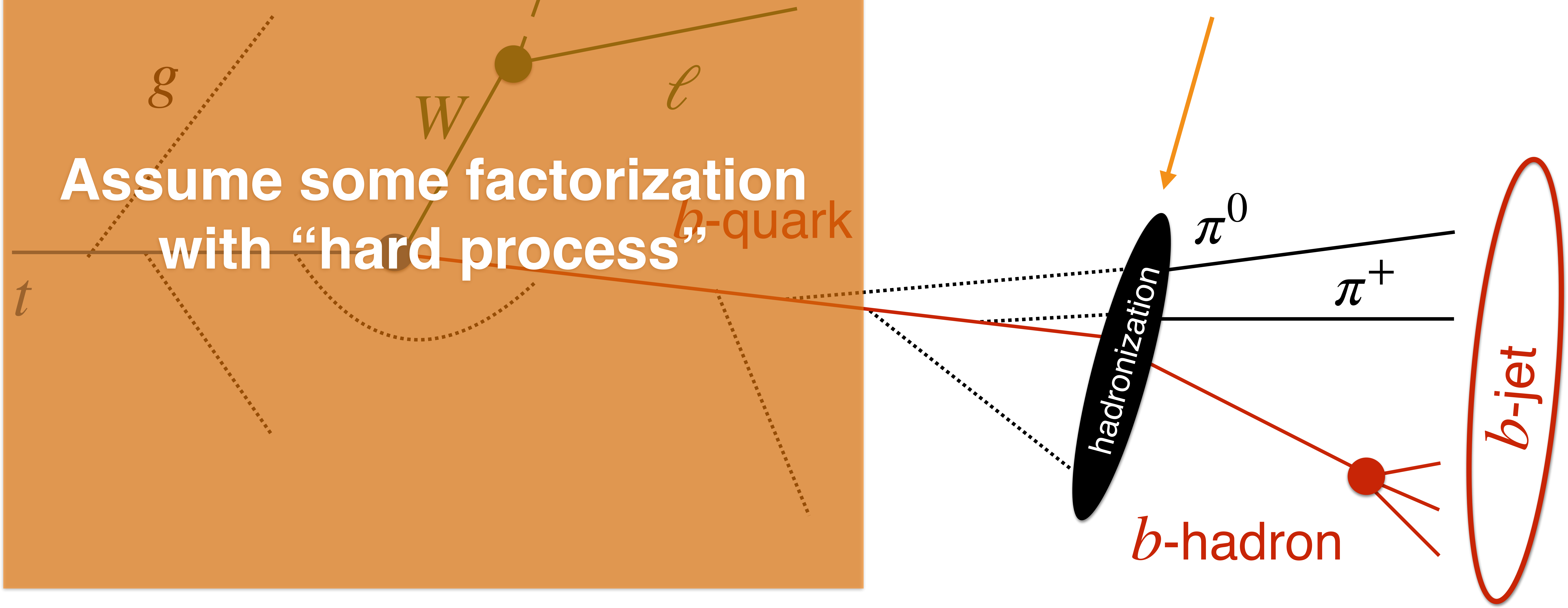
parton-shower generators are also not in good agreement

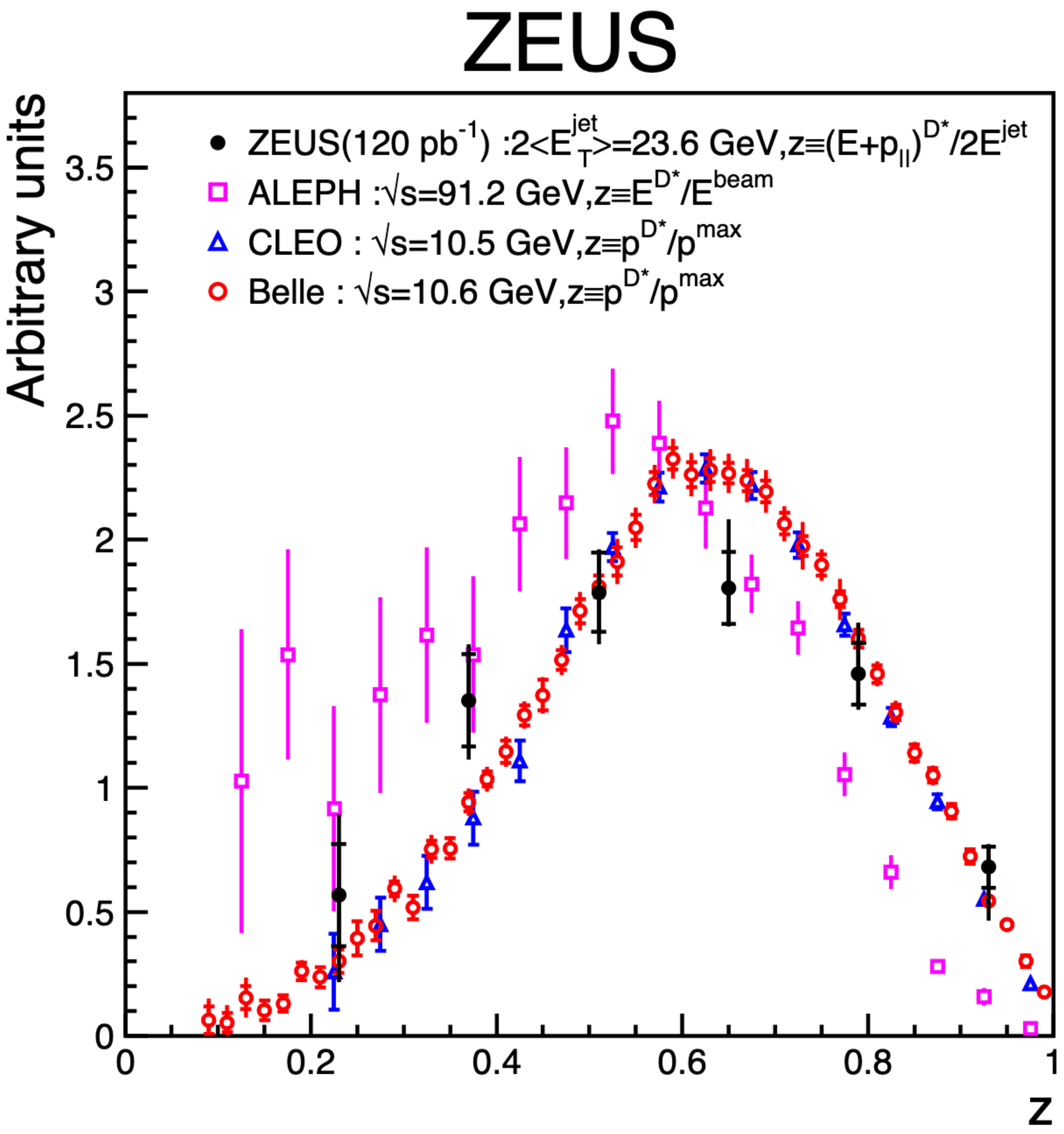
- b- and c-hadrons leave a striking experimental signature
- there is a unique correspondence to the originating heavy flavor quarks
    - ergo a precise probe of QCD
- b-fragmentation currently tuned to e^+e^- data from $Z \to bb$ decays.
- … then extrapolated to the LHC environment
    - to what degree is this correct?



tension between $e^+e^- \to bb$ measurements of b-fragmentation

parton-shower generators are also not in good agreement for a calculation with some accuracy in α_S, measure e.g. the b-hadron momentum fraction and apply universally





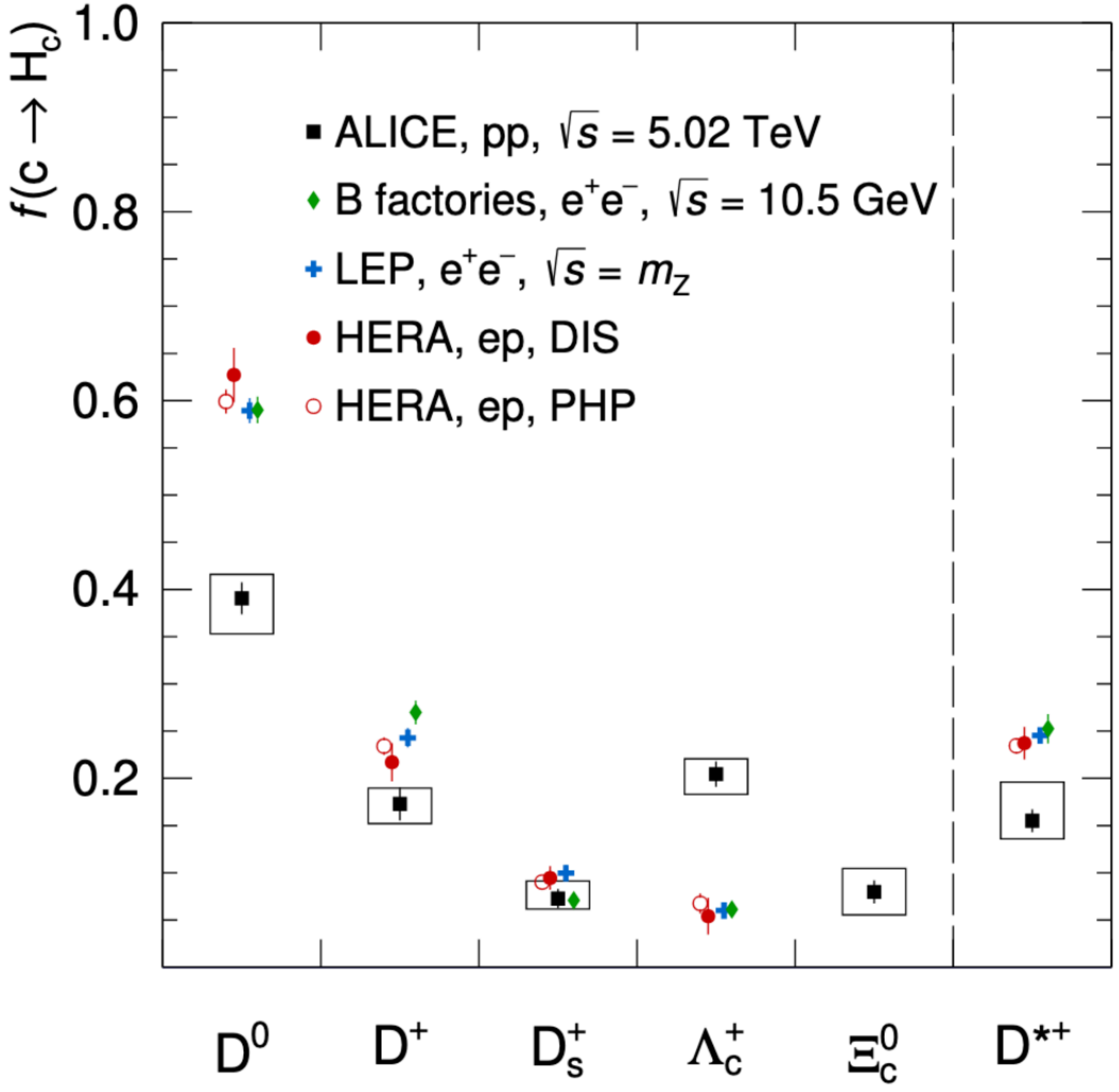
measurements of charm hadron momentum fractions are much less precise or understood.

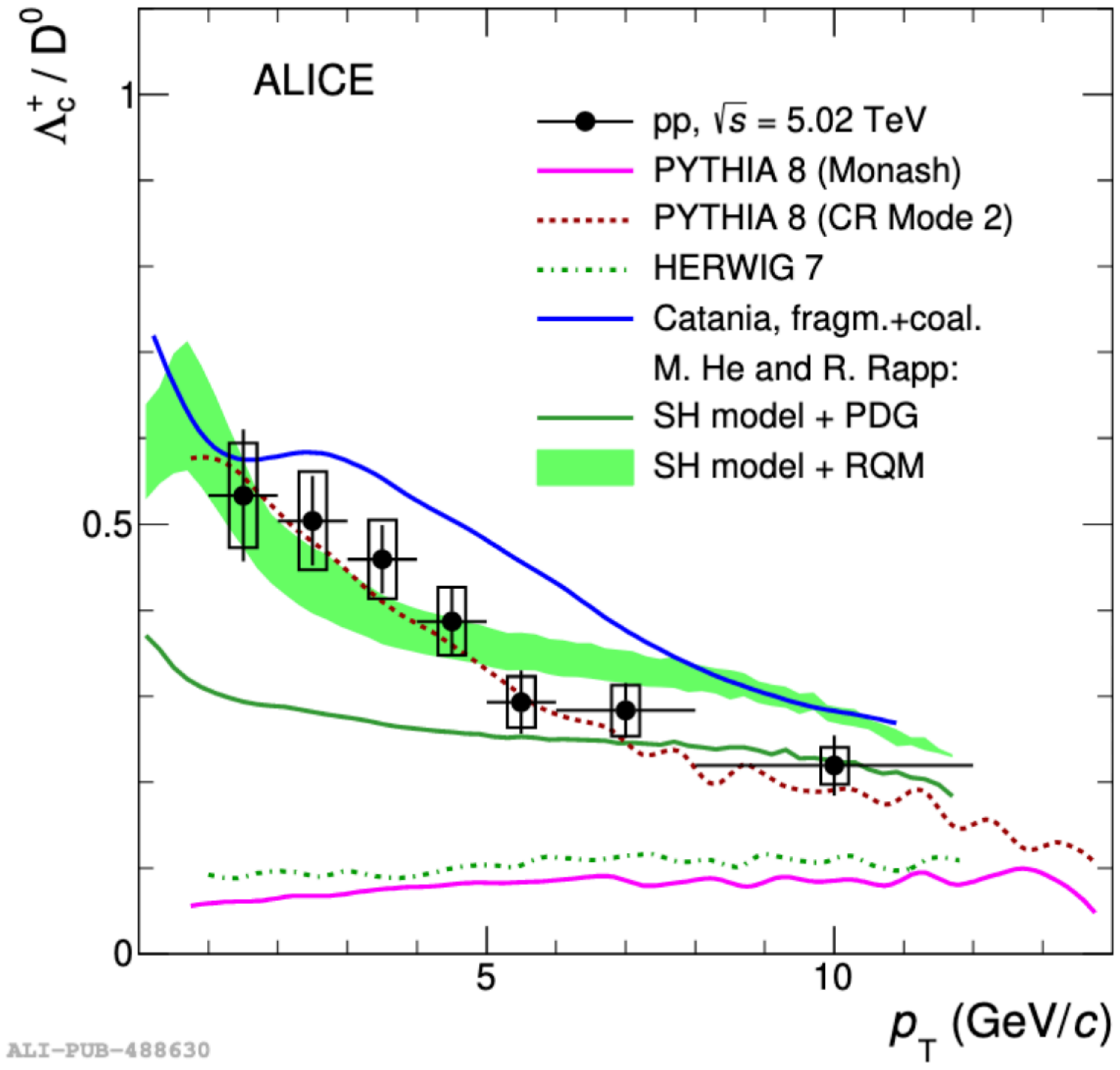
to my knowledge they are not currently used in any MC generator tunes.

(indeed, the Pythia Monash paper asks for more comprehensive measurements…)

ALICE are building an impressive body of work in the charm sector.

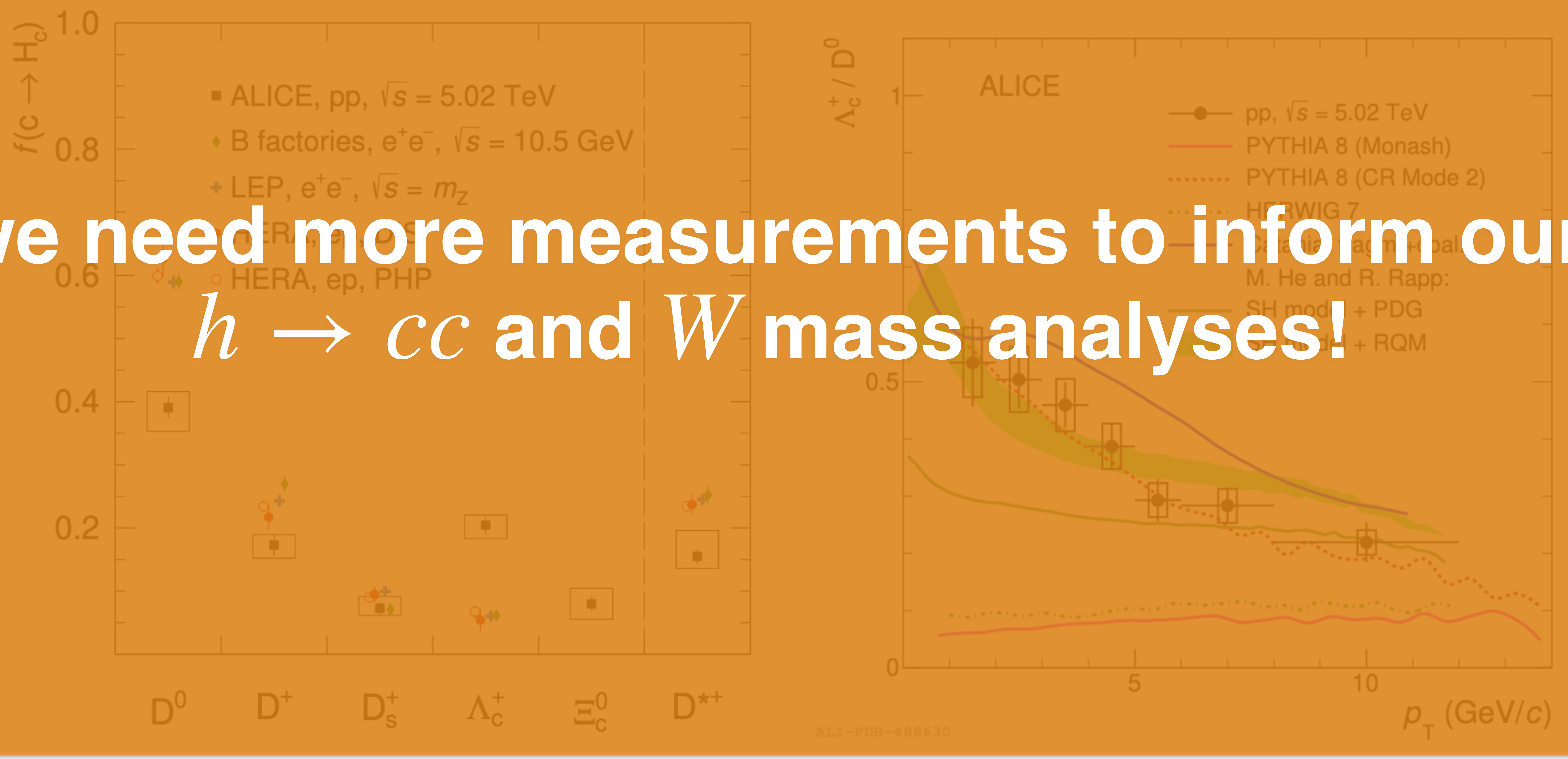
showing non-universality of charmed hadron production



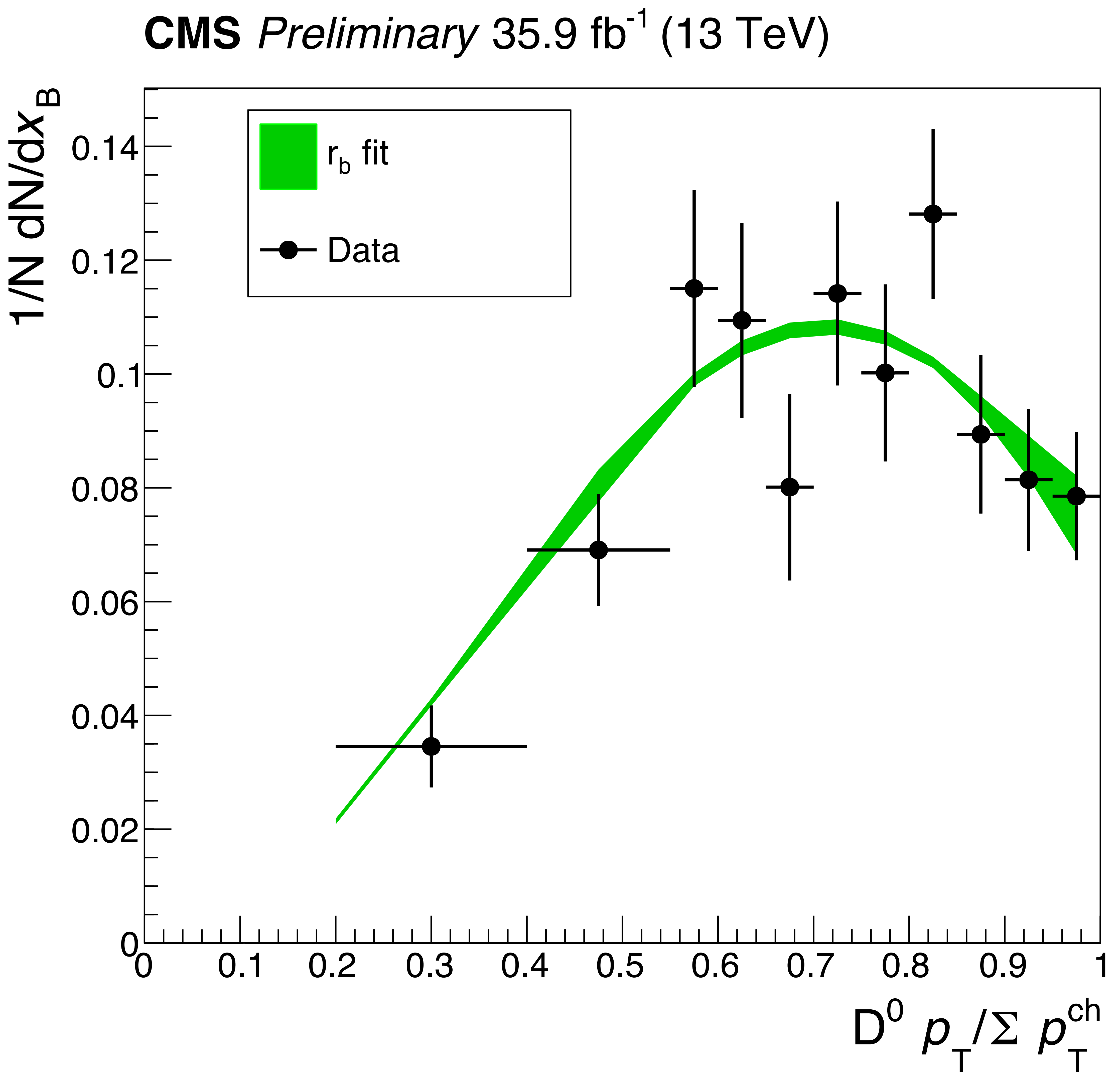


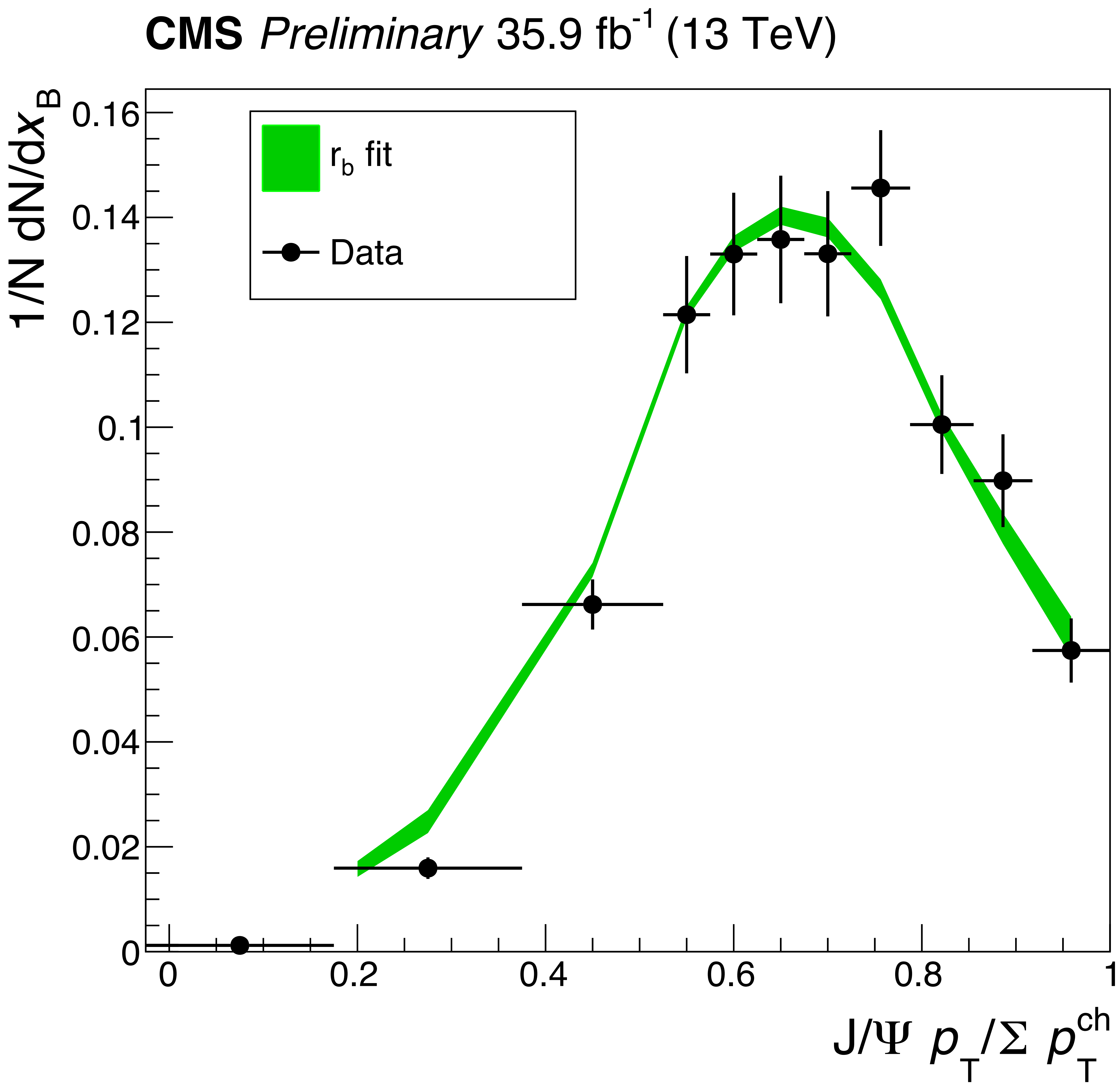
ALICE are building an impressive body of work in the charm sector.

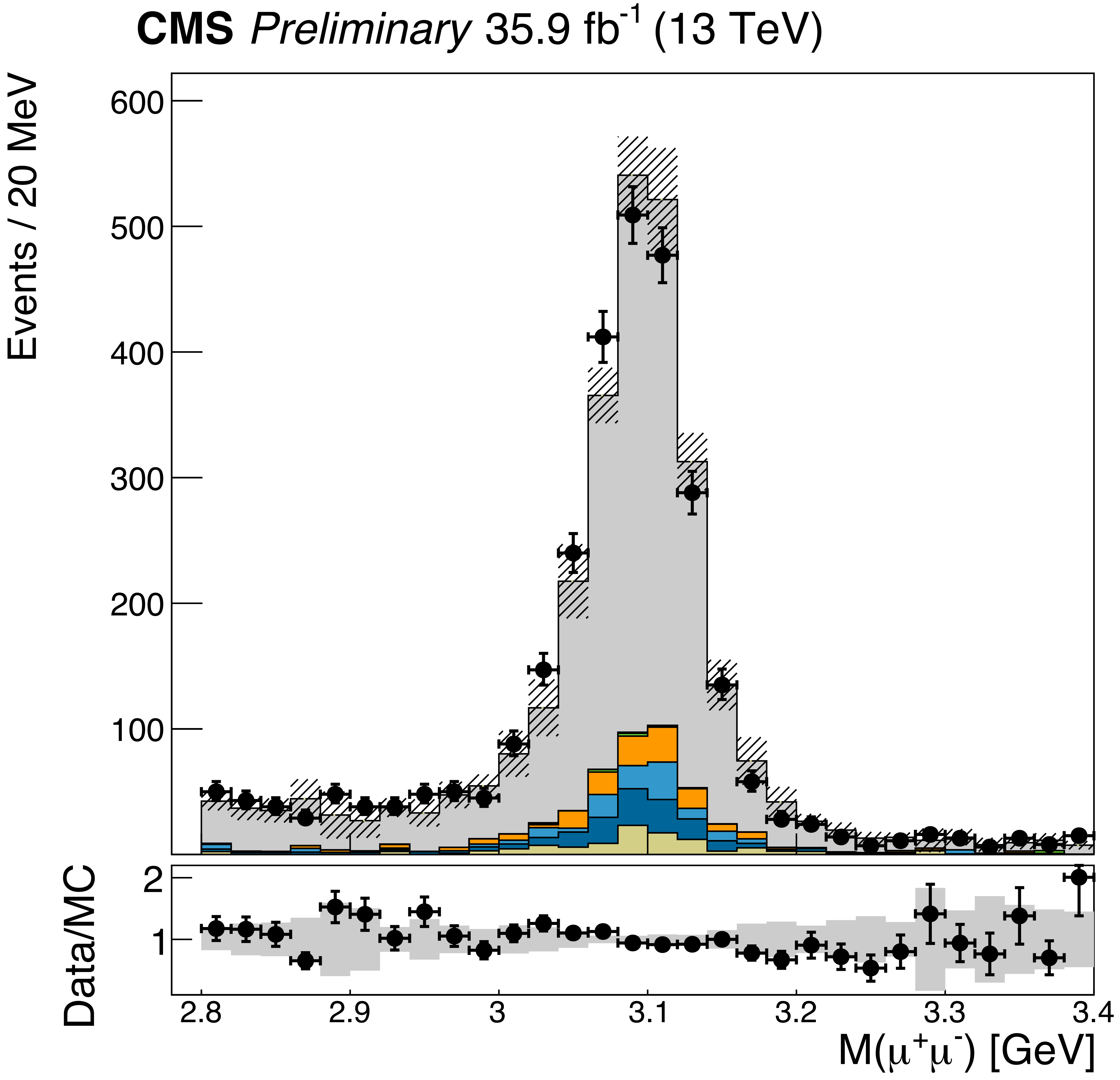
showing non-universality of charmed hadron production



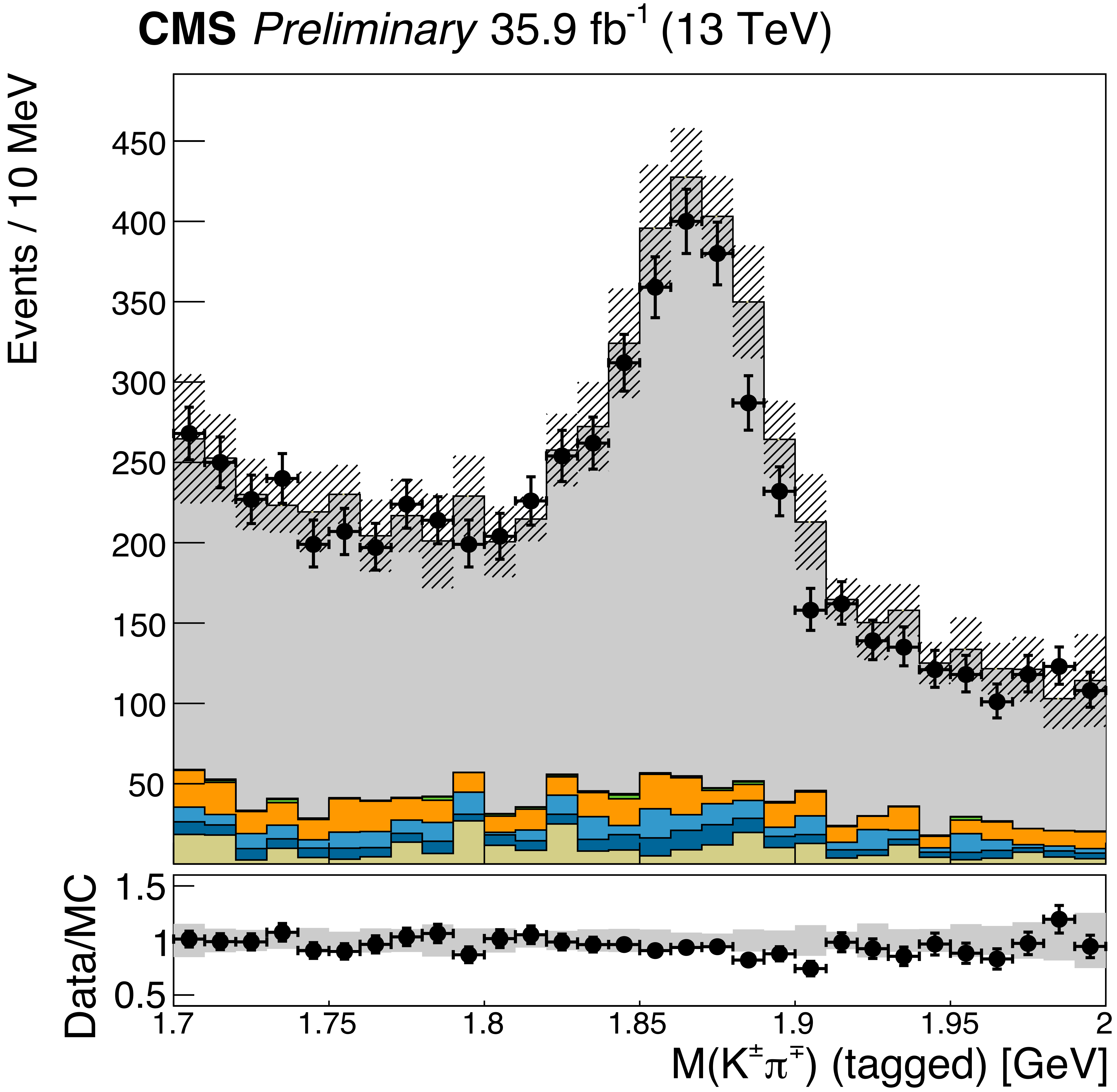
we need more measurements to inform our $h \to cc$ and W mass analyses!


CMS Preliminary 35.9 fb-1 (13 TeV)
1/N dN/dx_B
r_b fit
Data
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
D0 p_T/Σ p_T^ch


CMS Preliminary 35.9 fb-1 (13 TeV)
1/N dN/dx_B
r_b fit
Data
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0 0.2 0.4 0.6 0.8 1
J/Ψ p_T/Σ p_T^ch



CMS Preliminary 35.9 fb-1 (13 TeV)
Events / 20 MeV
Data/MC
M(μ+μ-) [GeV]



CMS Preliminary 35.9 fb-1 (13 TeV)
Events / 10 MeV
Data/MC
M(K±π∓) (tagged) [GeV]