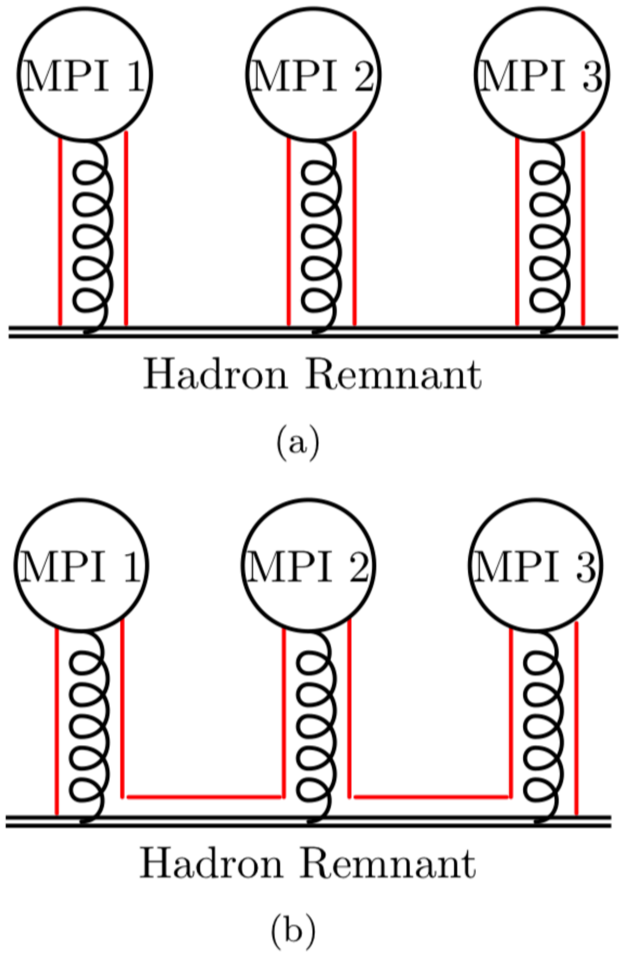
Extraction and validation of CMS PYTHIA 8 color reconnection tunes

Sercan SEN

(on behalf of the CMS Collaboration)

What is Color Reconnection (CR)?



Christiansen, J.R., Skands, P.Z. JHEP, 2015, 3

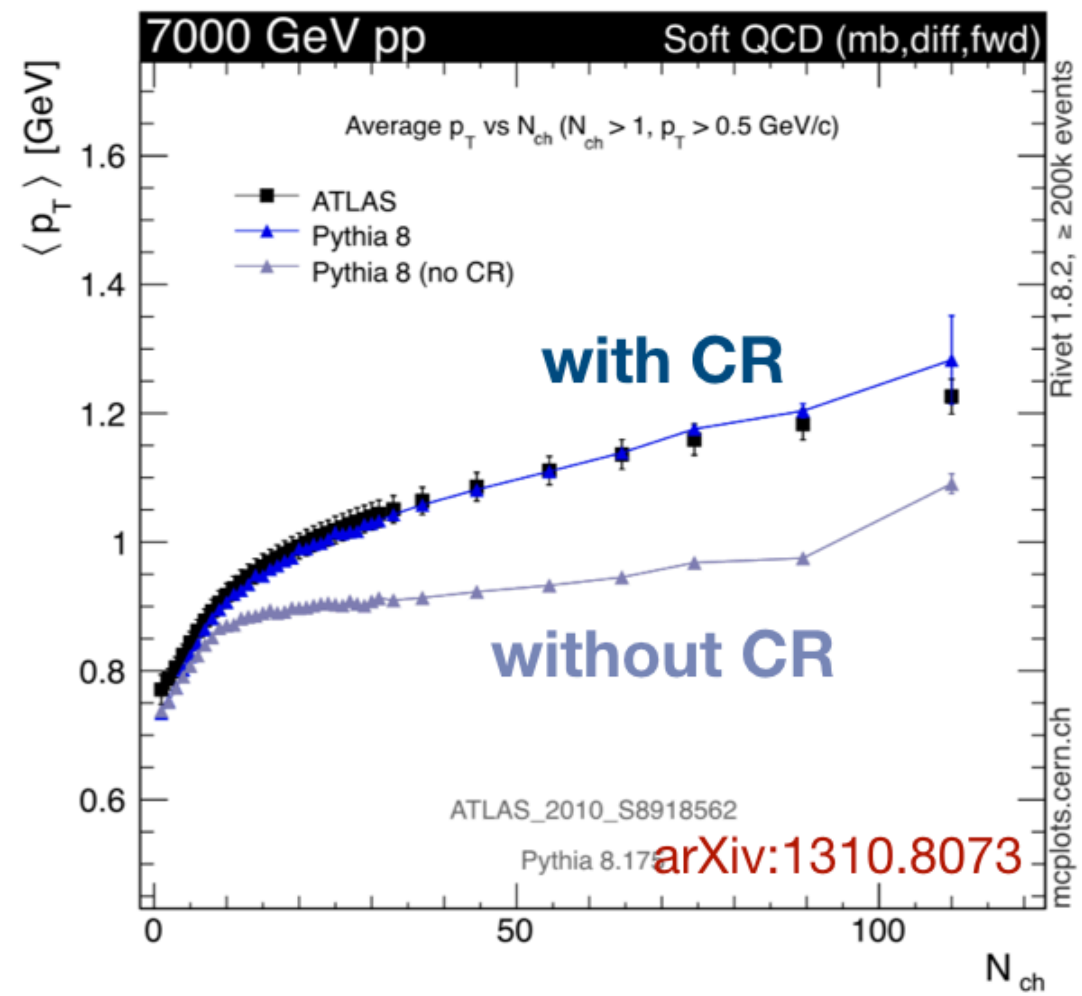
Leading colour (LC) approximation:

- Each MPI is viewed as separate from all other systems in colour space.
- No strings stretched between different MPI systems.

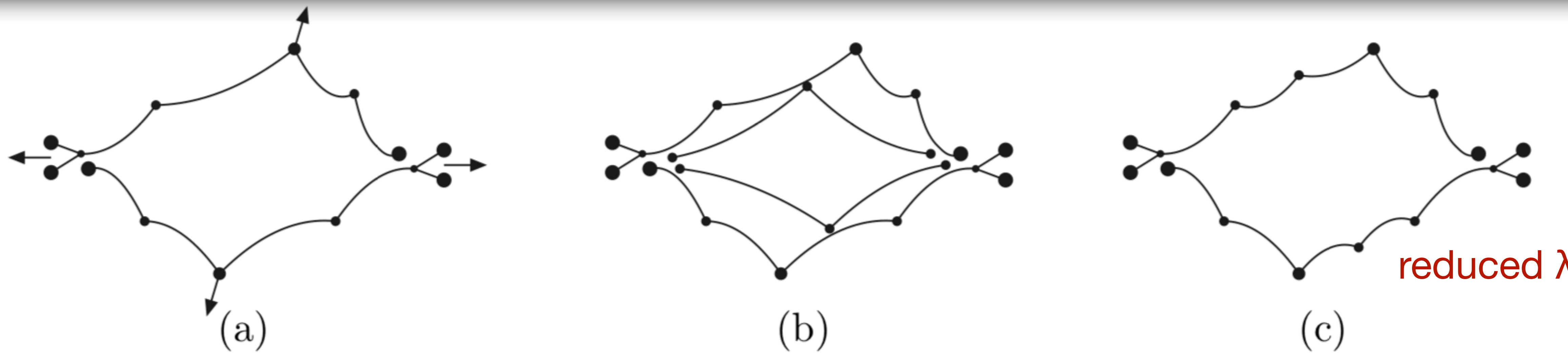
Color reconnection allows different MPI systems to be colour-connected to each other. MPI hadronize collectively. Total colour charge reduced w.r.t. LC approximation.

Rising trend of ⟨p⊥⟩(nch):

- first observed by UA1 [Nucl. Phys. B335 (1990)]
- Colour reconnection is needed to describe the data.



MPI-based CR model in PYTHIA8



First hard scaterring: outgoing gluons are colour connected

A second hard scattering: gives new strings connected to the remnants.

gluons are colour reconnected, so that the total string length (λ) becomes as short as possible.

Reconnection probability

$$P_{\mathrm{rec}}(p_{\mathrm{T}}) = \frac{(R_{\mathrm{rec}} \cdot p_{\mathrm{T0}})^2}{R_{\mathrm{rec}} \cdot p\mathrm{T0} + p_{\mathrm{T}}^2}$$

ColourReconnection:range

(free parameter of the model)

The higher this number is the more reconnections can occur.

Reduce λ by adding partons of the lower-pT system to the strings defined by the higher-pT system

$$\lambda \approx \sum_{i,j} \ln\left(\frac{m_{ij}^2}{m_0^2}\right)$$

$p_{\mathrm{T}} \downarrow \Rightarrow P_{\mathrm{rec}} \uparrow$

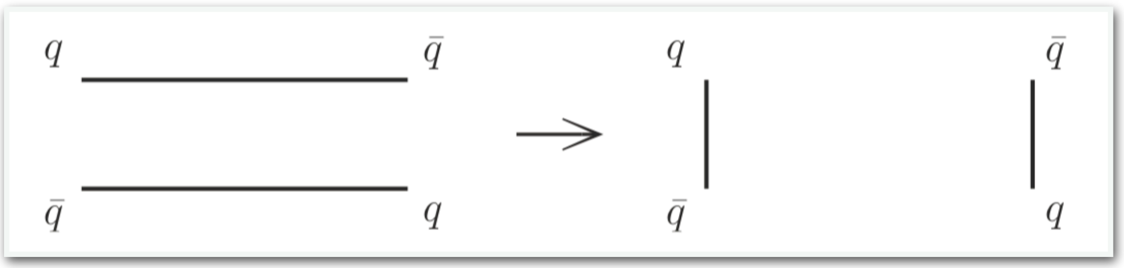
softer systems easier to reconnect

G. Gustafson, Acta Phys. Polon. B40, 1981 (2009)

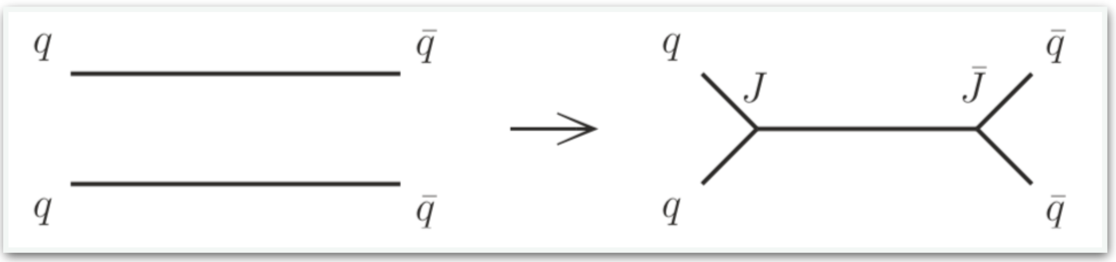
QCD-based CR model in PYTHIA8

Another approach for reconnection ⇒ introduce "junctions"

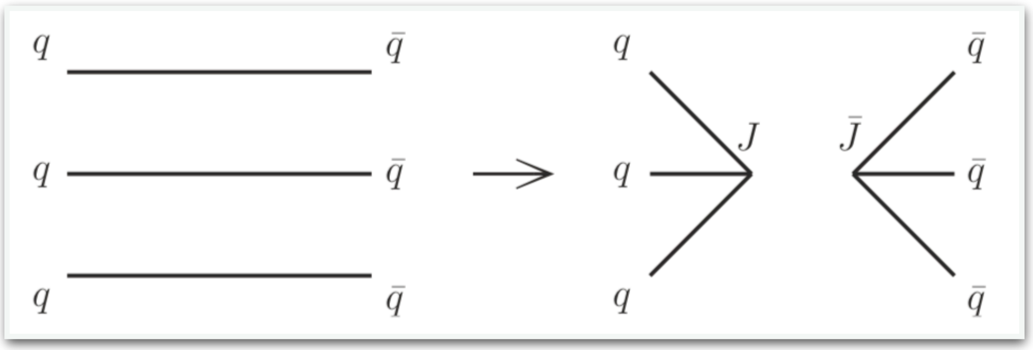
- More realistic; colour topologies defined by the SU(3) colour algebra.
- Improves description of baryon production.



ordinary string reconnection



double junction reconnection



triple junction reconnection

more types of junction reconnections available..

Some free parameters of the model:

m0: variable used in the λ measure for the string length.

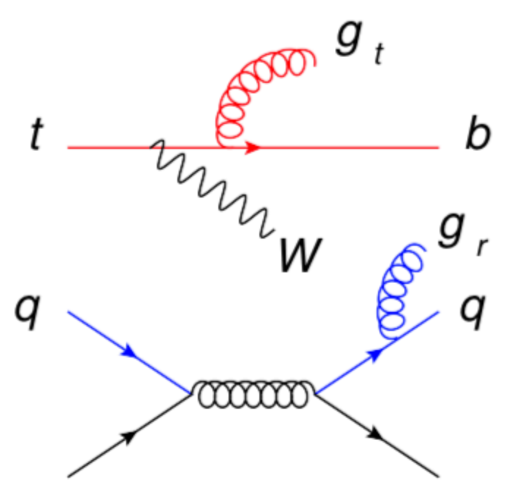
timeDilationPar: controls the time of two strings to resolve each other between formation and hadronization.

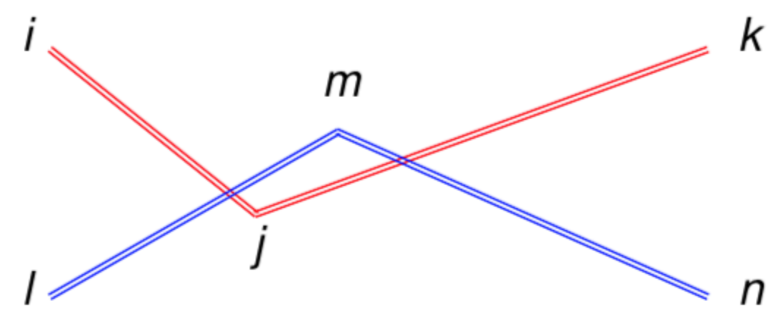
junctionCorrection: multiplicative factor to control junction production.

Christiansen, J.R., Skands, P.Z. JHEP, 2015, 3

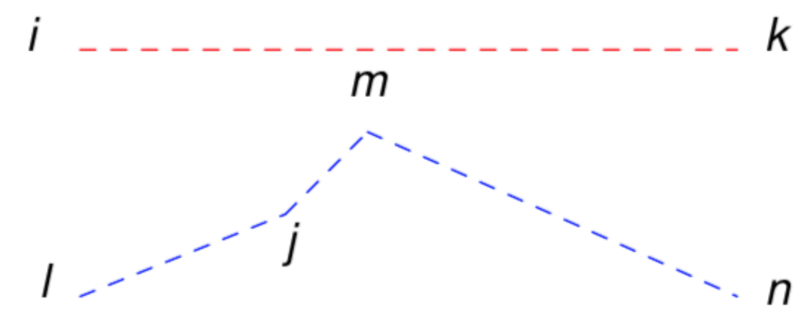
gluon-move CR model in PYTHIA8

Similar to other CR models, it aims to reduce the total "string length" λ.





original configuration



resulting configuration after moving the gluon (reduced λ)

→ A gluon *j* originally attached to a string piece, *ik*, can be moved to a different string piece, *lm*, if it leads to a smaller total string length λ.

$$\Delta\lambda(j,lm) = \lambda_{j;lm} - \lambda_{j;ik} = \lambda_{lj} + \lambda_{jm} + \lambda_{ik} - (\lambda_{ij} + \lambda_{jk} + \lambda_{lm})\,.$$

$$\min_{j,lm} \Delta\lambda(j,lm) \leq \Delta\lambda_{\mathrm{cut}}$$

Some free parameters of the model:

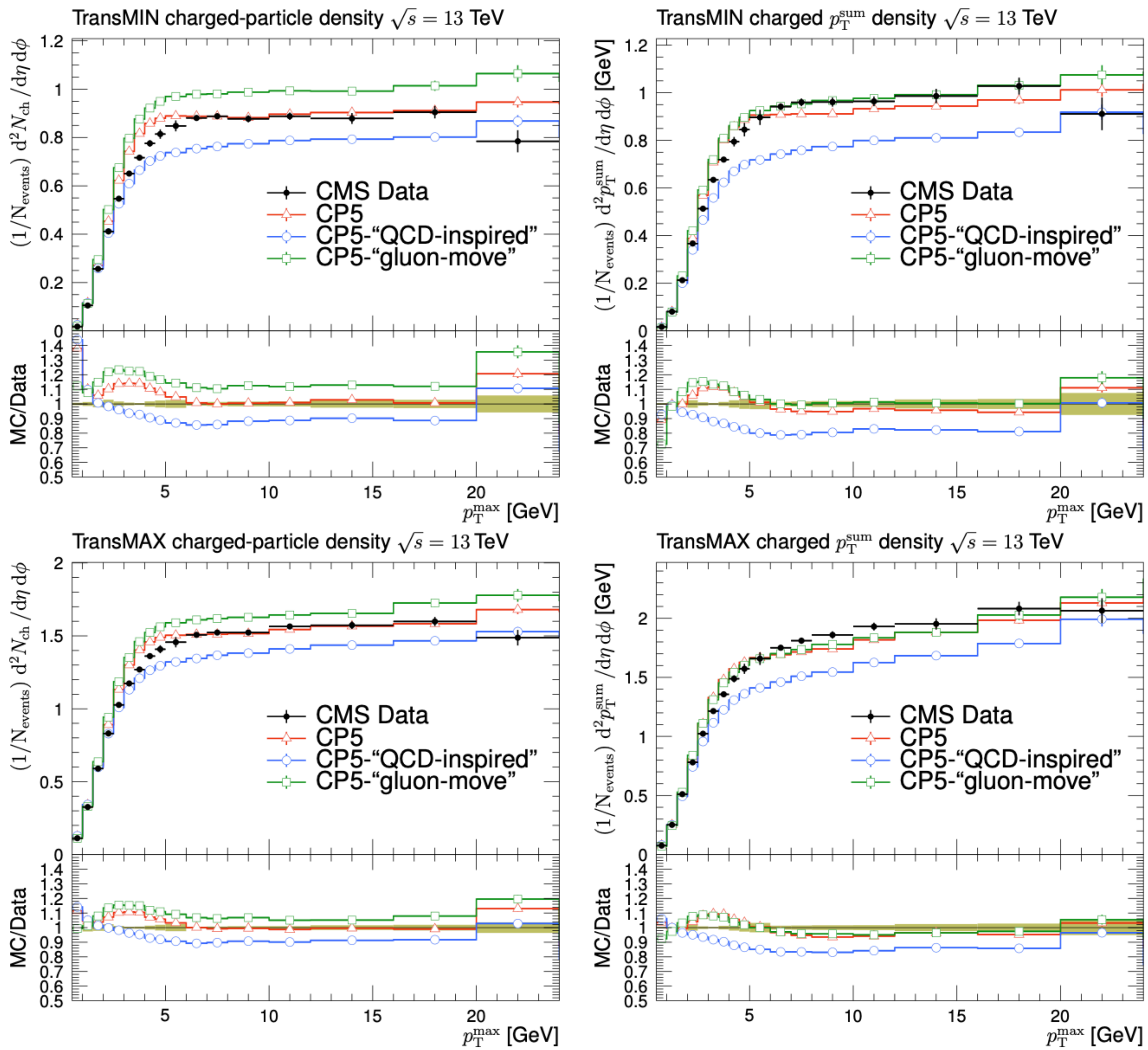
m2Lambda: variable used in the definition of λ.

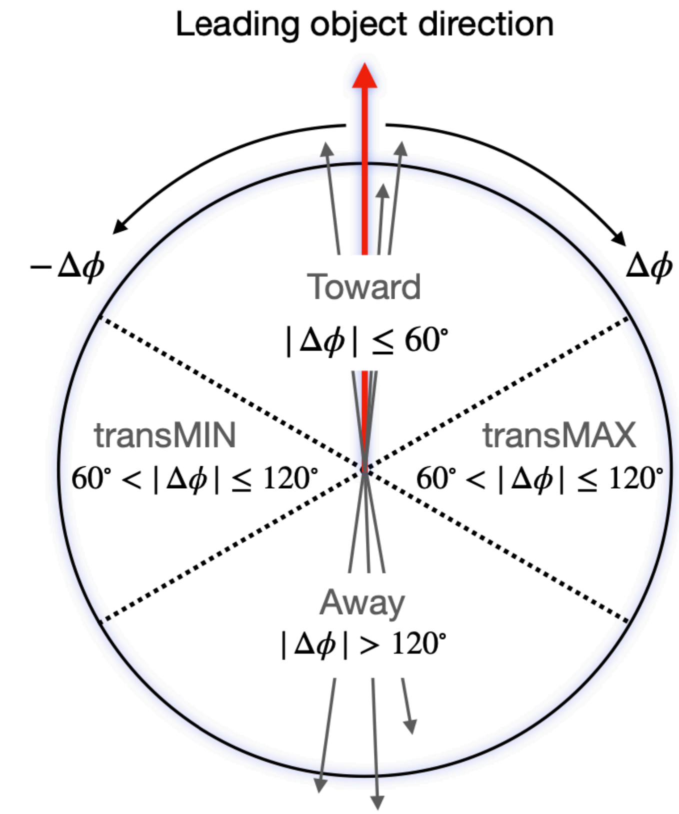
fracGluon: probability that a given gluon will be considered for being moved.

Argyropoulos, S., Sjöstrand, T., JHEP, 2014, 43

Model predictions for MB/UE observables at 13 TeV

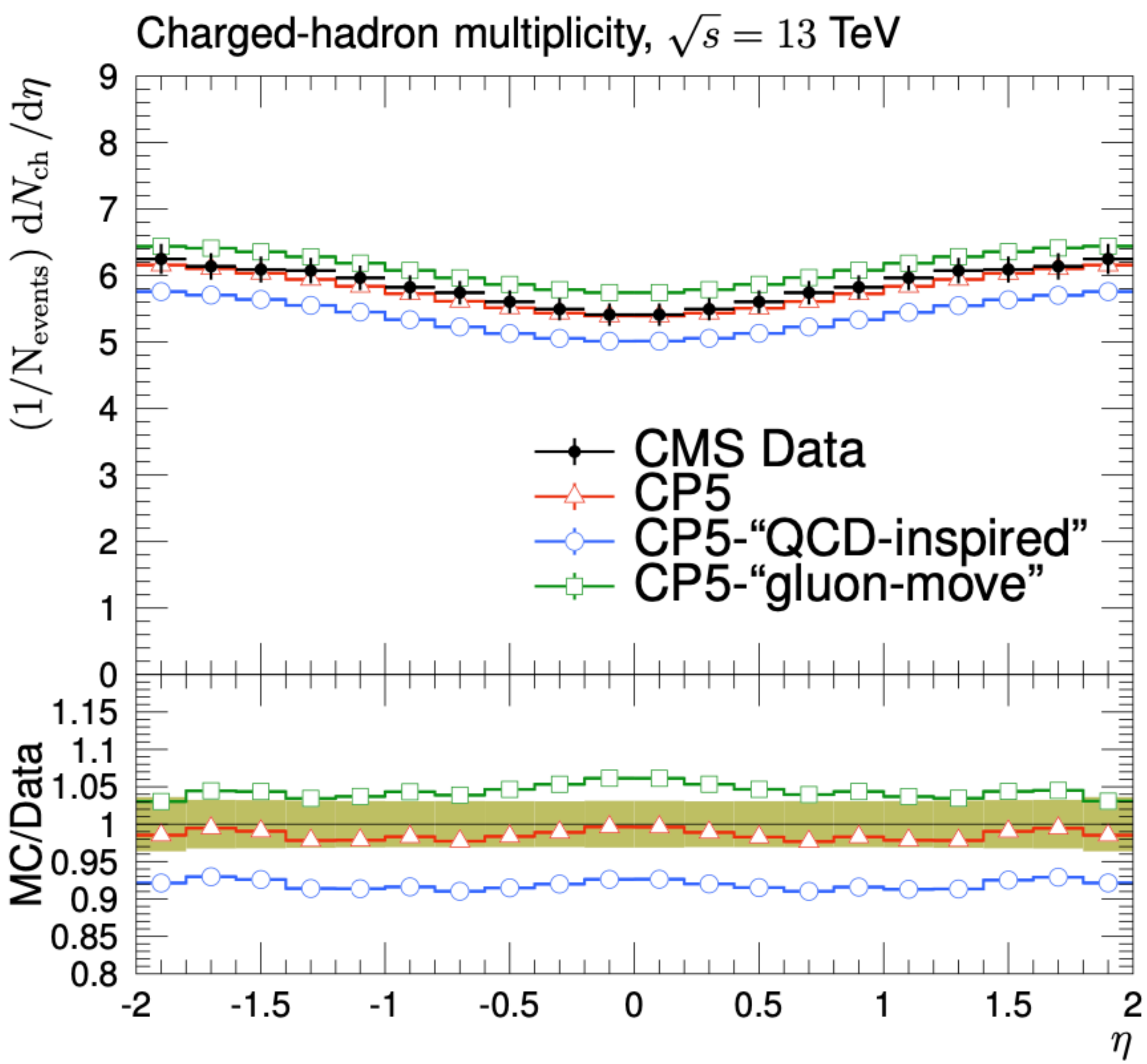
Charged particle and ptsum densities in TransMin and TransMax





- CR models with their default parameter values in PYTHIA do not describe UE data at 13 TeV.
- Gluon-move model is in good agreement with p_T^{sum} data, but 10-20% overestimates the charged particle density.
- QCD-inspired model underestimates every distribution.

Model predictions for MB/UE observables at 13 TeV



- QCD-inspired model produces less charged hadrons.
- Gluon-move model describes the data within uncertainties, but there is still room for tuning.

Models must be tuned to describe the underlying soft physics in pp collisions at 13 TeV.

Tuning procedure

- **Framework**: PYTHIA 8.226 + Rivet 2.4.0 + Professor 1.4.0
- **Models**: **QCD-inspired (CP5-CR1)** and **gluon-move (CP5-CR2)** CR models in PYTHIA8.
- **Baseline**: CP5 tune with NNPDF3.1_nnlo_as_0118 PDF set, 0.118 for ISR, FSR, and MPI.
- **Input observables/rivet routines*:**

				CP5-CR1		CP5-CR2	
RIVET routine	$\sqrt{s}$ [TeV]	$\lvert\eta\rvert$	Distribution	Fit range [GeV]	R	Fit range [GeV]	R
CMS_2015_I1384119	13	<2.0	N_{ch} versus η		1		1
CMS_2015_PAS_FSQ_15_007	13	<2.0	TransMIN charged p_T^{sum}	2–28	1	3–36	0.5
			TransMAX charged p_T^{sum}	2–28	1	3–36	0.5
			TransMIN N_{ch}	2–28	1	3–36	0.1
			TransMAX N_{ch}	2–28	1	3–36	0.1
CMS_2012_PAS_FSQ_12_020	7	<0.8	TransMAX N_{ch}	3–20	1	3–20	0.1
			TransMIN N_{ch}	3–20	1	3–20	0.1
			TransMAX charged p_T^{sum}	3–20	1	3–20	0.1
			TransMIN charged p_T^{sum}	3–20	1	3–20	0.1
CDF_2015_I1388868	2	<0.8	TransMIN N_{ch}	2–15	1	2–15	0.1
			TransMAX N_{ch}	2–15	1	2–15	0.1
			TransMIN charged p_T^{sum}	2–15	1	2–15	0.1
			TransMAX charged p_T^{sum}	2–15	1	2–15	0.1

* The fit ranges and R are slightly different for CP1

R: relative importance

Tuning procedure

- Generate predictions at ~200 points in parameter space and parametrize
- For each parameter set pp inelastic events are simulated.
- Professor interpolates the bin values for the observables in the parameter space using a 3rd-order polynomial (order of the polynomial does not affect the results significantly).
- **Tune to data by minimizing:**

models the MC response of each bin b of the observable O as a function of the parameter vector p

data value for each bin b

$$\chi^{*2}(\mathbf{p}) = \sum_{O} \sum_{b \in O} \frac{(f^b(\mathbf{p}) - \mathcal{R}_b)^2}{\Delta_b^2}$$

Total bin uncertainty of the data

Not a true χ^2 function which treats all distributions/bins equally (i.e., relative importance; $R = 1$). Poor description of dN_{ch}/dη with CP5-CR2 when $R = 1$ for all distributions. However, dN_{ch}/dη is a key observable and increasing its relative importance in the fit is reasonable.

Tune results

fixed parameters

tuned parameters (CR + MPI)

PYTHIA 8 parameter	CP5 [10]	CP5-CR1	CP5-CR2
PDF set	NNPDF3.1 NNLO	NNPDF3.1 NNLO	NNPDF3.1 NNLO
$\alpha_S(m_Z)$	0.118	0.118	0.118
`SpaceShower:rapidityOrder`	on	on	on
`MultipartonInteractions:ecmRef` [GeV]	7000	7000	7000
$\alpha_S^{ISR}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{FSR}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{MPI}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{ME}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
`StringZ:aLund`	—	0.38	—
`StringZ:bLund`	—	0.64	—
`StringFlav:probQQtoQ`	—	0.078	—
`StringFlav:probStoUD`	—	0.2	—
`SigmaTotal:zeroAXB`	off	off	off
`BeamRemnants:remnantMode`	—	1	—
`ColourReconnection:mode`	—	1	2
`MultipartonInteractions:pT0Ref` [GeV]	1.410	1.375	1.454
`MultipartonInteractions:ecmPow`	0.033	0.033	0.054
`MultipartonInteractions:coreRadius`	0.763	0.605	0.649
`MultipartonInteractions:coreFraction`	0.630	0.445	0.489
`ColourReconnection:range`	5.176	—	—
`ColourReconnection:junctionCorrection`	—	0.238	—
`ColourReconnection:timeDilationPar`	—	8.580	—
`ColourReconnection:m0`	—	1.721	—
`ColourReconnection:m2lambda`	—	—	4.917
`ColourReconnection:fracGluon`	—	—	0.993
N_{dof}	183	157	158
χ^{*2}/N_{dof}	1.04	2.37	0.89

- For CR1, string fragmentation parameters proposed in [1] are used in the initial settings.
- Similar to CP5, CR tunes prefer a "colder core" profile: radius > fraction.
- fracGluon parameter in CR2 prefers a value very close to 1. Almost all gluons are taken into account to be moved.

[1] Christiansen, J.R., Skands, P.Z. JHEP, 2015, 3

Tune Performances

Underlying event and minimum-bias observables (1.96, 7, 8, and 13 TeV)

- N_{ch} and charged $p_{\mathrm{T}}^{\mathrm{sum}}$ densities in TransMin and TransMax regions,
- Mean average p_{T} in TransMin and TransMax regions,
- Charged-hadron pseudorapidity distribution
- Charged particle pseudorapidity distribution
- Charged particle and charged $p_{\mathrm{T}}^{\mathrm{sum}}$ in transverse region
- Energy density in the forward region (for inelastic and also for non-single diffractive events)

Particle multiplicities (ee colliders, 7 and 13 TeV)

- **Strange particle production (7 TeV)**
  - Λ baryons and $\mathrm{K}_{\mathrm{S}}^{0}$ rapidity distribution
- **Identified particles**
  - p/π ratio as a function of p_{T}
  - Baryon/meson ratio (for light, charm, and bottom hadrons)

☑ **New tunes are tested in various final states and center-of-mass energies**

Jet substructure observables (7 TeV)

- Average charged-hadron multiplicity as a function of jet p_{T}
- fragmentation function for jets

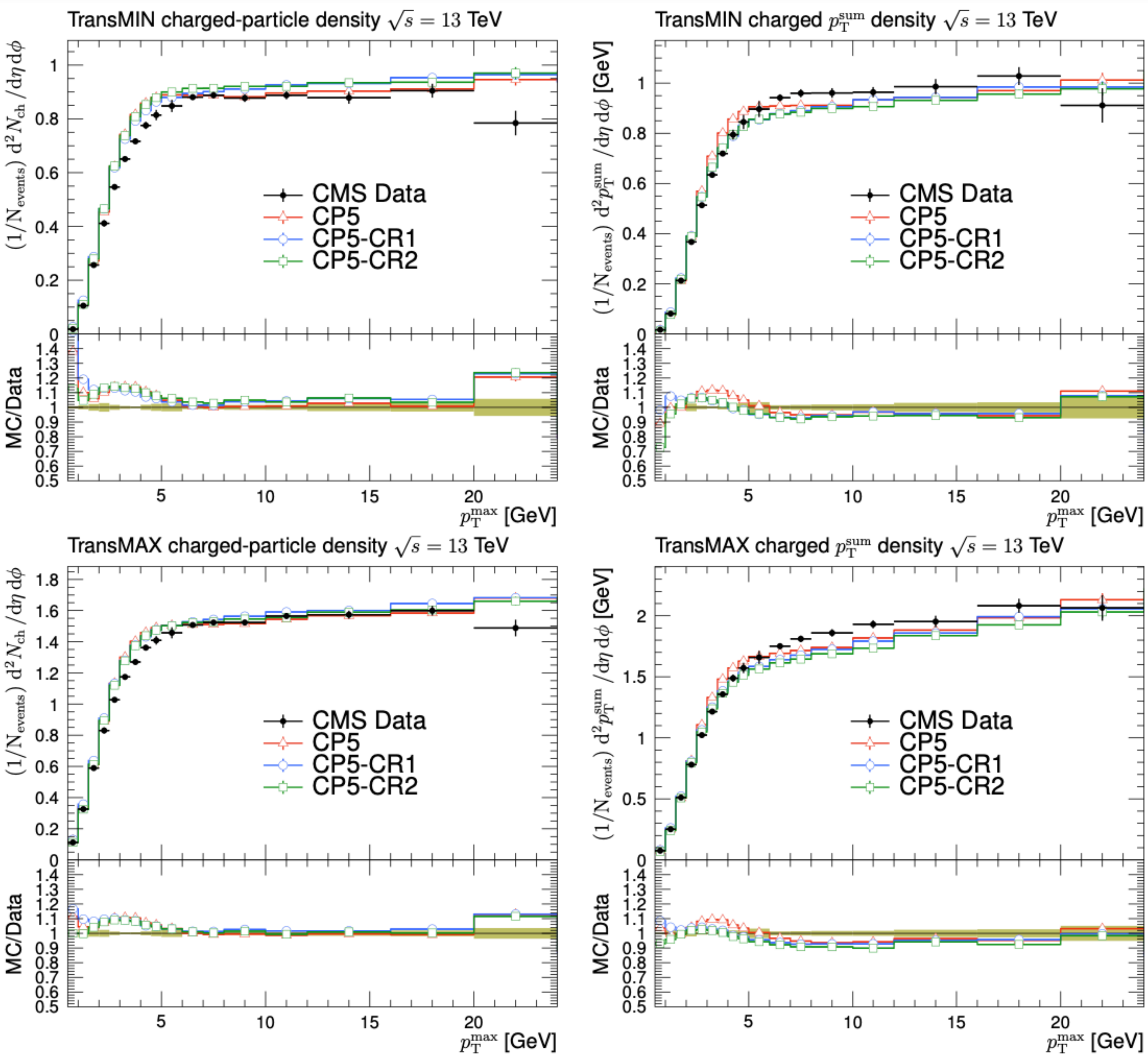
Drell-Yan events (13 TeV)

- Number of charged particles and p_{T} flow in the transverse region

Top quark observables (8 and 13 TeV)

- Jet substructure in $t\bar{t}$ events (charged particle multiplicity in jets and angle between two jets)
- Jet pull angle in $t\bar{t}$ events

Tune Performances: MB/UE at 13 TeV



•Good description of plateau by all tunes. The low p_T part (≈5 GeV of p_T^{max}) highly sensitive to diffractive contributions. Below ≈2-3 GeV is not included in the fits.

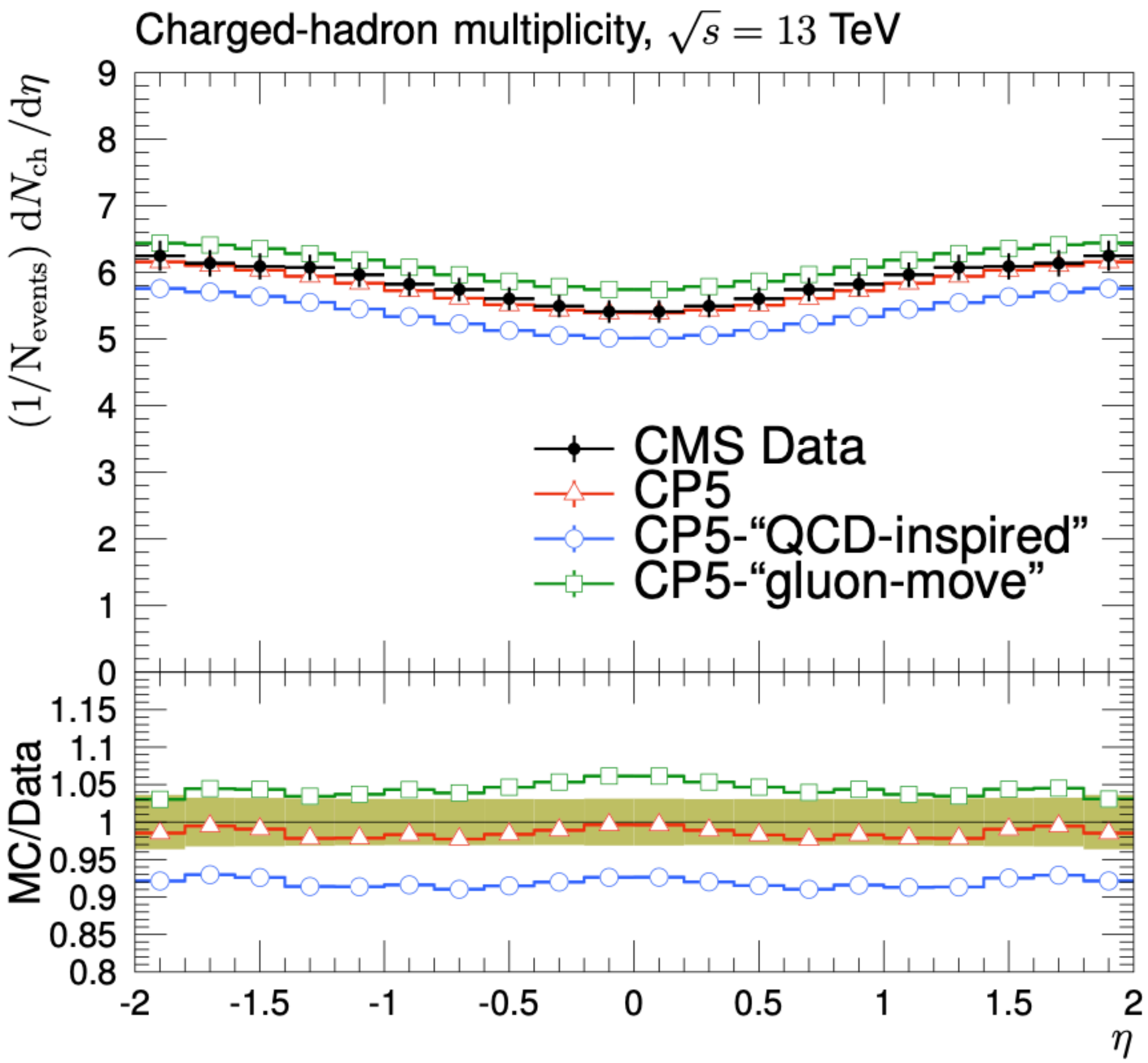
•New tunes perform better compared to tunes using default parameters of the new CR models.

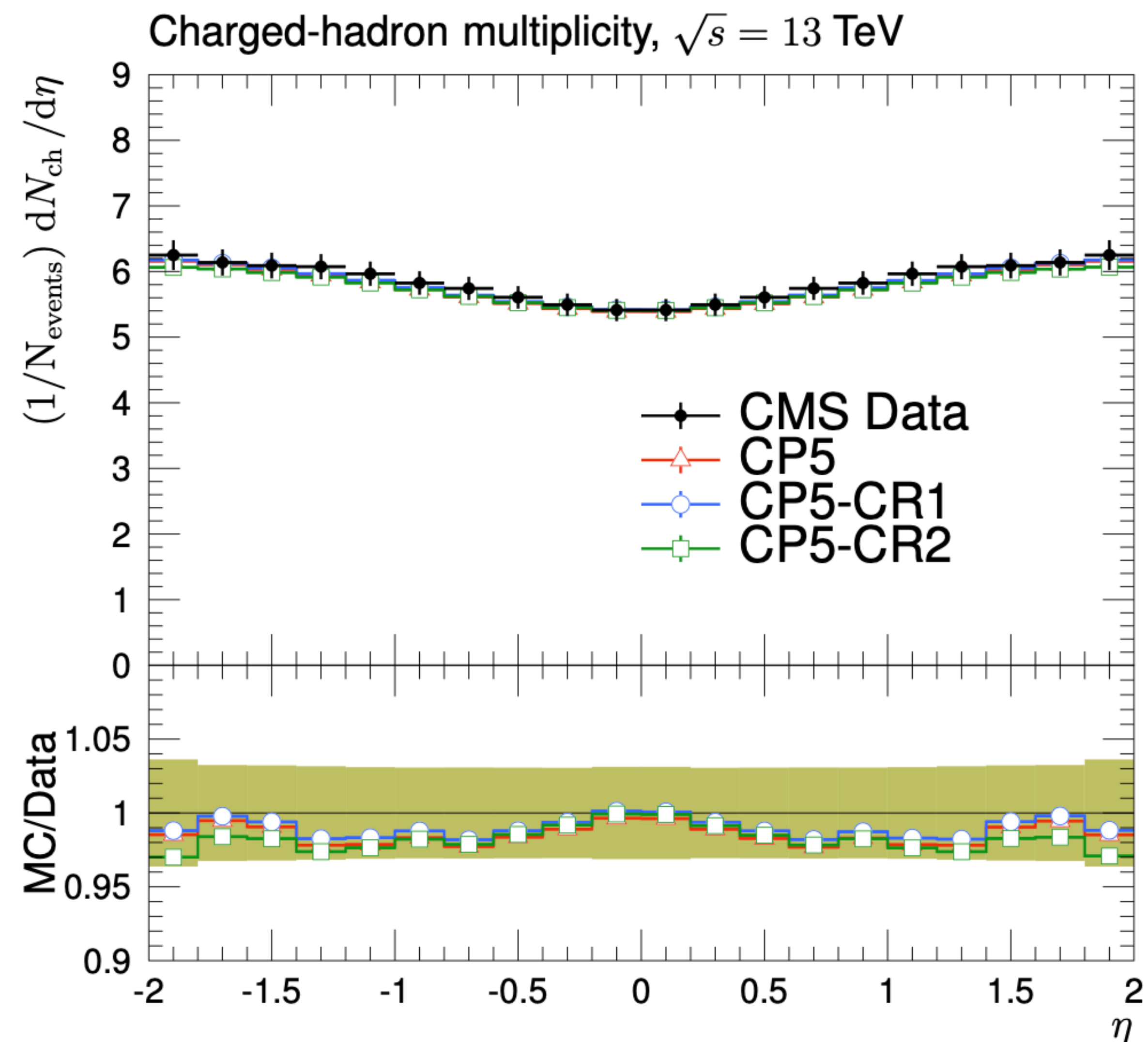
•Predictions of the new CR tunes are similar to that of CP5, but no better than CP5.

•Reasonable agreement with the data also for MB/UE at lower center-of-mass energies.

Tune Performances: MinBias at 13 TeV

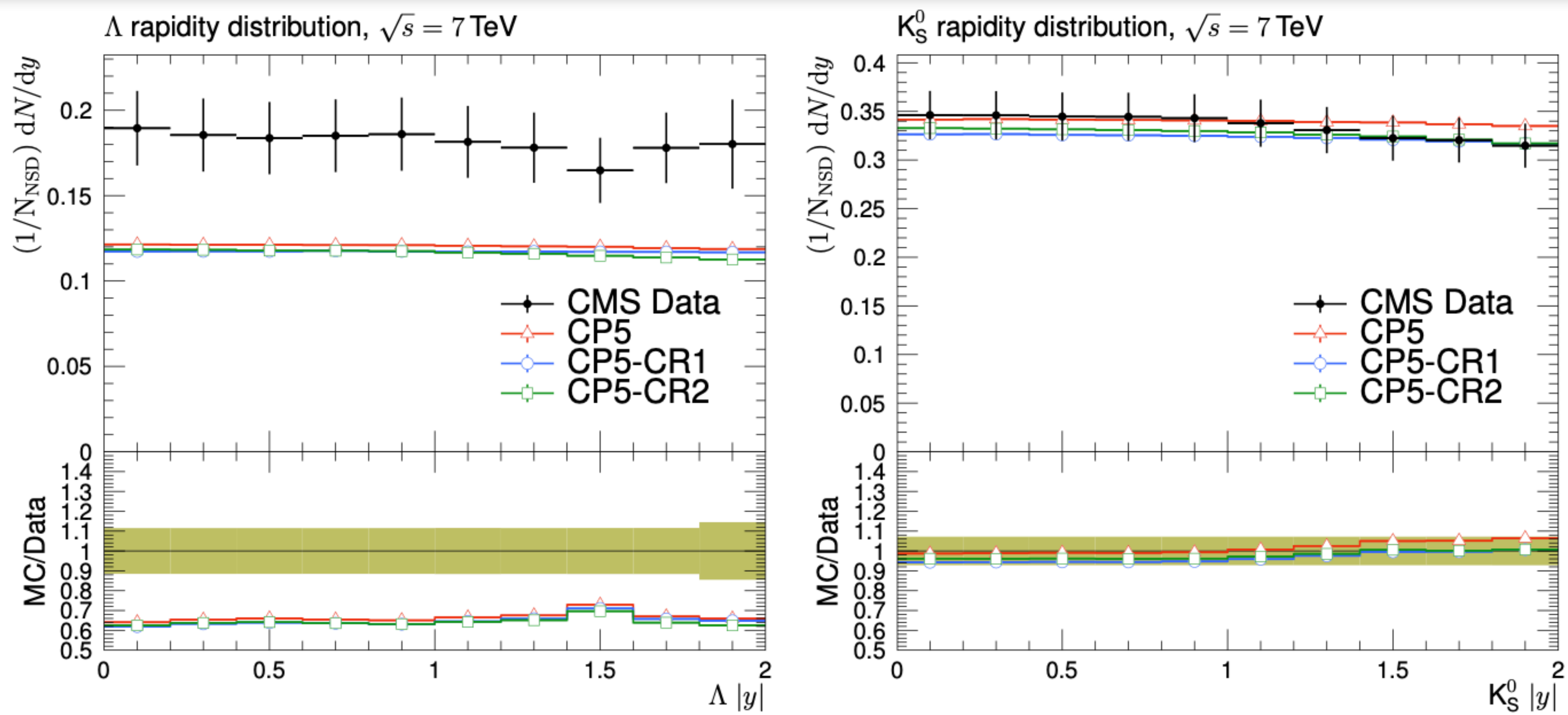
After tuning





- Key observable sensitive to the softer part of the MPI spectrum. Perfectly described by all tunes (CP5, CP5-CR1, and CP5-CR2)
- Predictions of tunes are similar for MB observables.

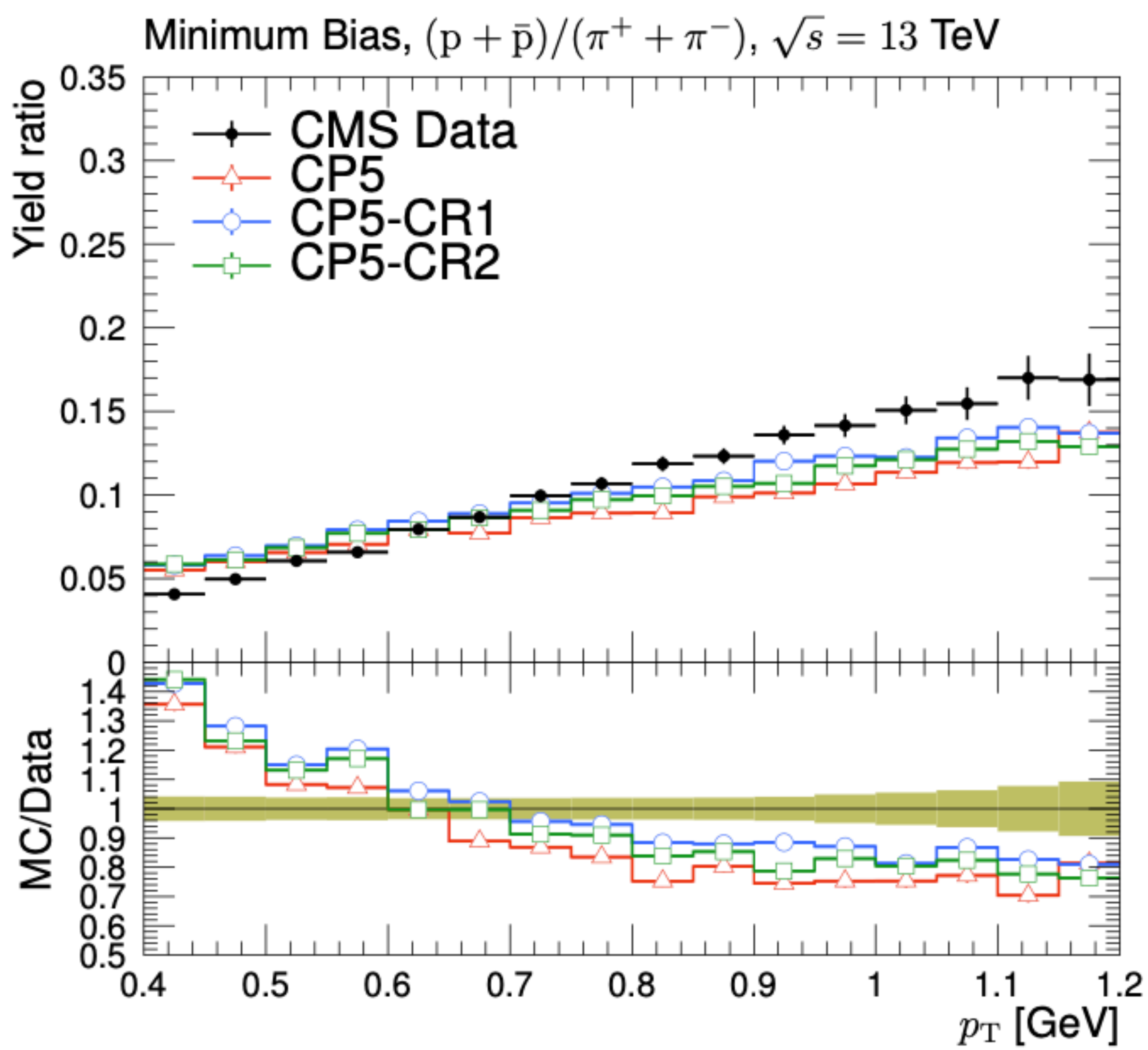
Tune Performances: Strange particle productions (7 TeV)

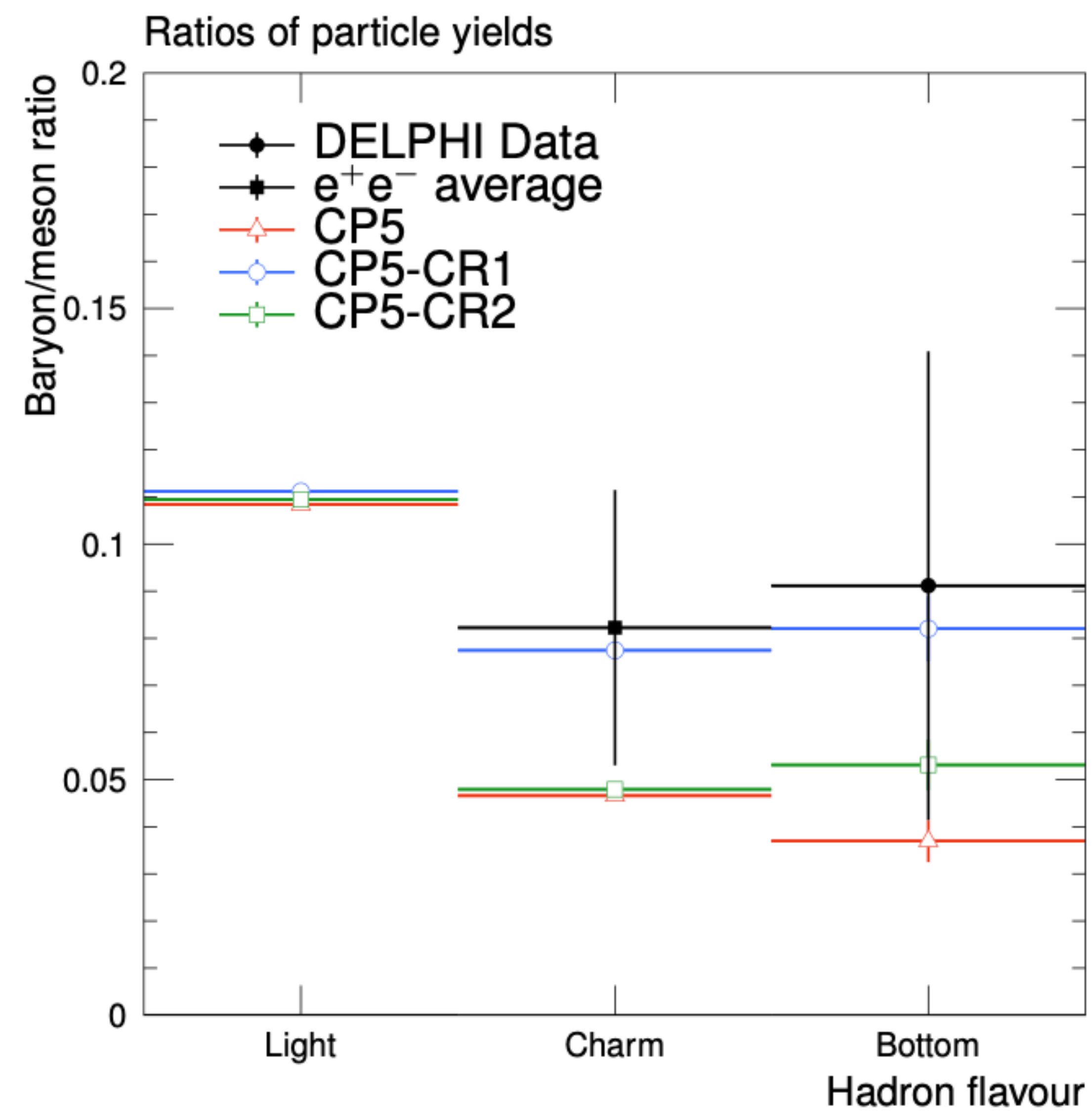


- It is a known issue that the LHC tunes do not perfectly describe this data.
  - Shown in "Christiansen, J.R., Skands, P.Z. JHEP, 2015, 3" that new QCD-based CR model improves Λ/K^0_S versus rapidity production in pp collisions. However, the study is based on Monash tune with a different PDF set, alphas and MPI values from CP5. (Monash: NNPDF2.3 LO PDF set). Also, data at 13 TeV was not included in the tuning.

- After our tuning to existing data:
  - All CP5 tunes, regardless of the CR model, describe K^0_S versus rapidity very well.
  - Λ versus rapidity is underestimated by 30%. Therefore, Λ/K^0_S is not perfectly described.

Tune Performances: Identified particles





p/π as a function of particle p_T.

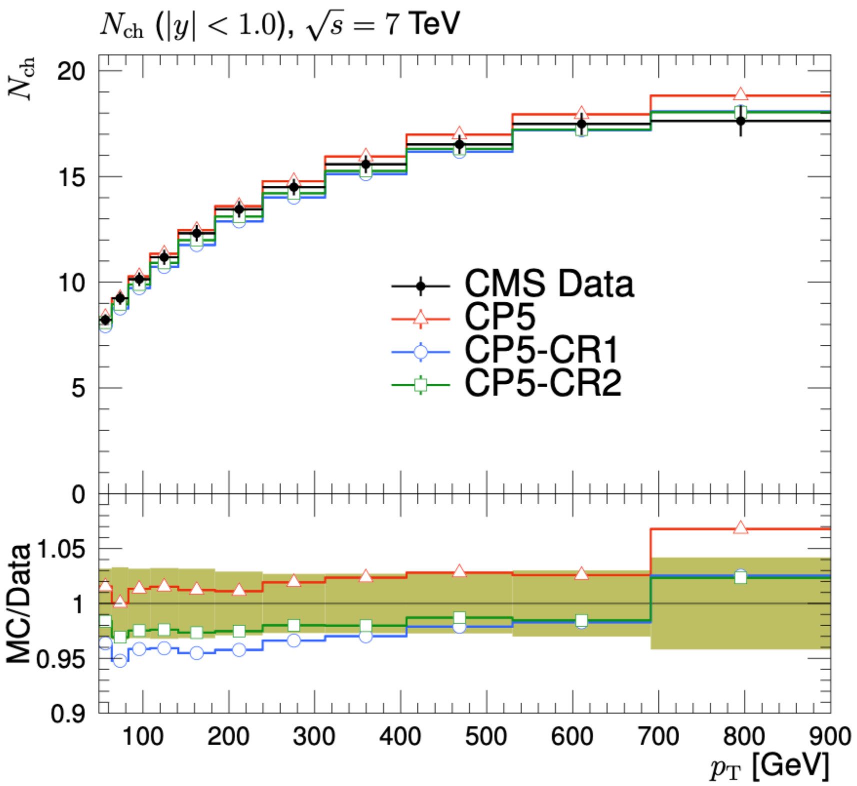
- New CR models do not improve the description of the data significantly. However, not only CR, but also MPI and hadronization are important.

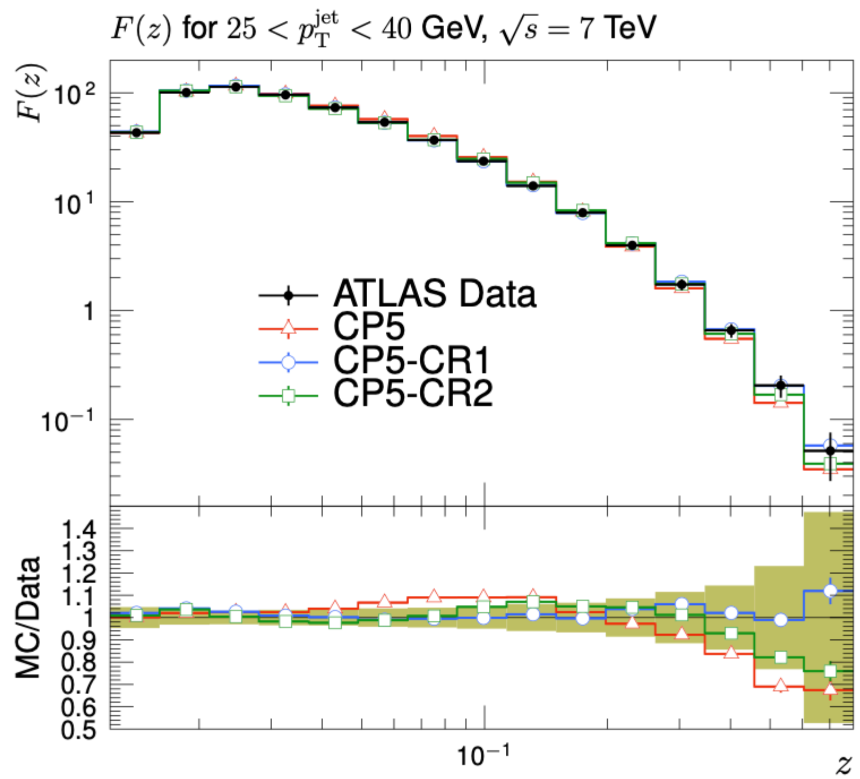
•Charm hadron data from e+e- colliders at $\sqrt{s} = \sim 10$ GeV and bottom hadron data from DELPHI at $\sqrt{s} = \sim 90$ GeV.

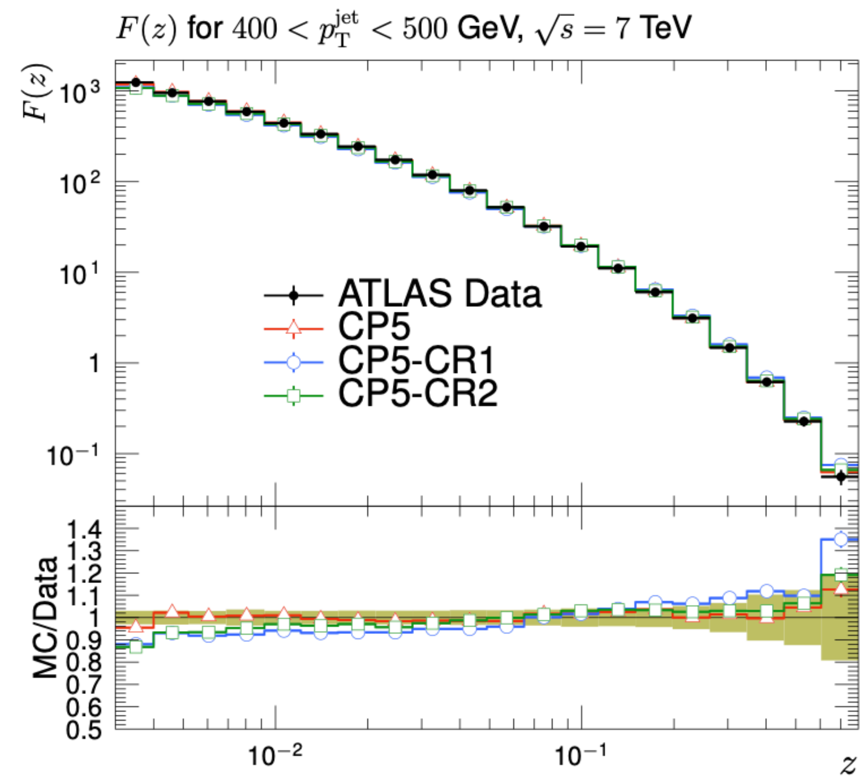
• Key observable for the correct particle spectra.

•Good description of baryon-to-meson yield ratios at different center-of-mass energies.

Tune Performances: Jet substructure observables (7 TeV)







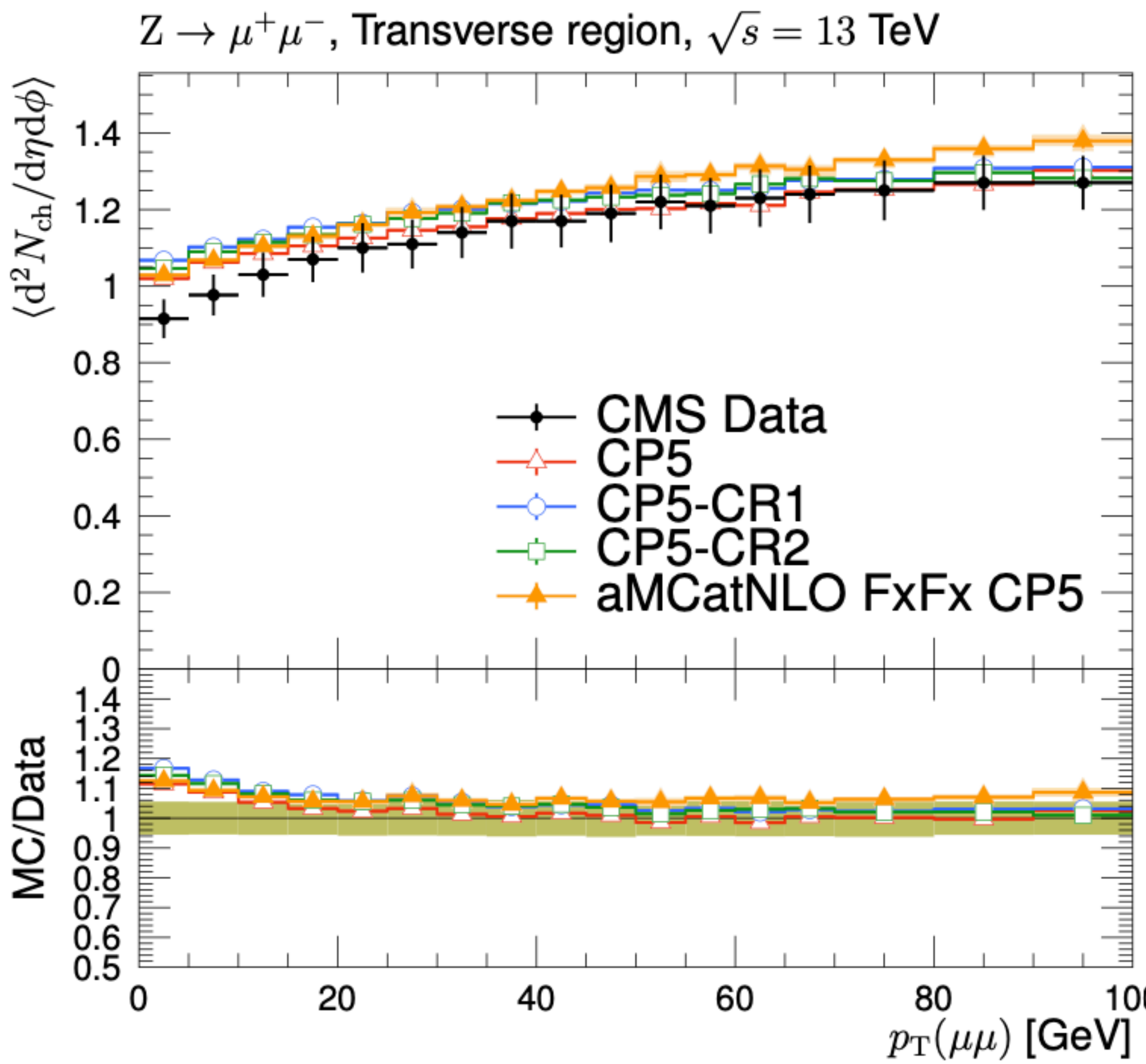
$$F(z) = (1/N_{\mathrm{jet}})(\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}z)$$

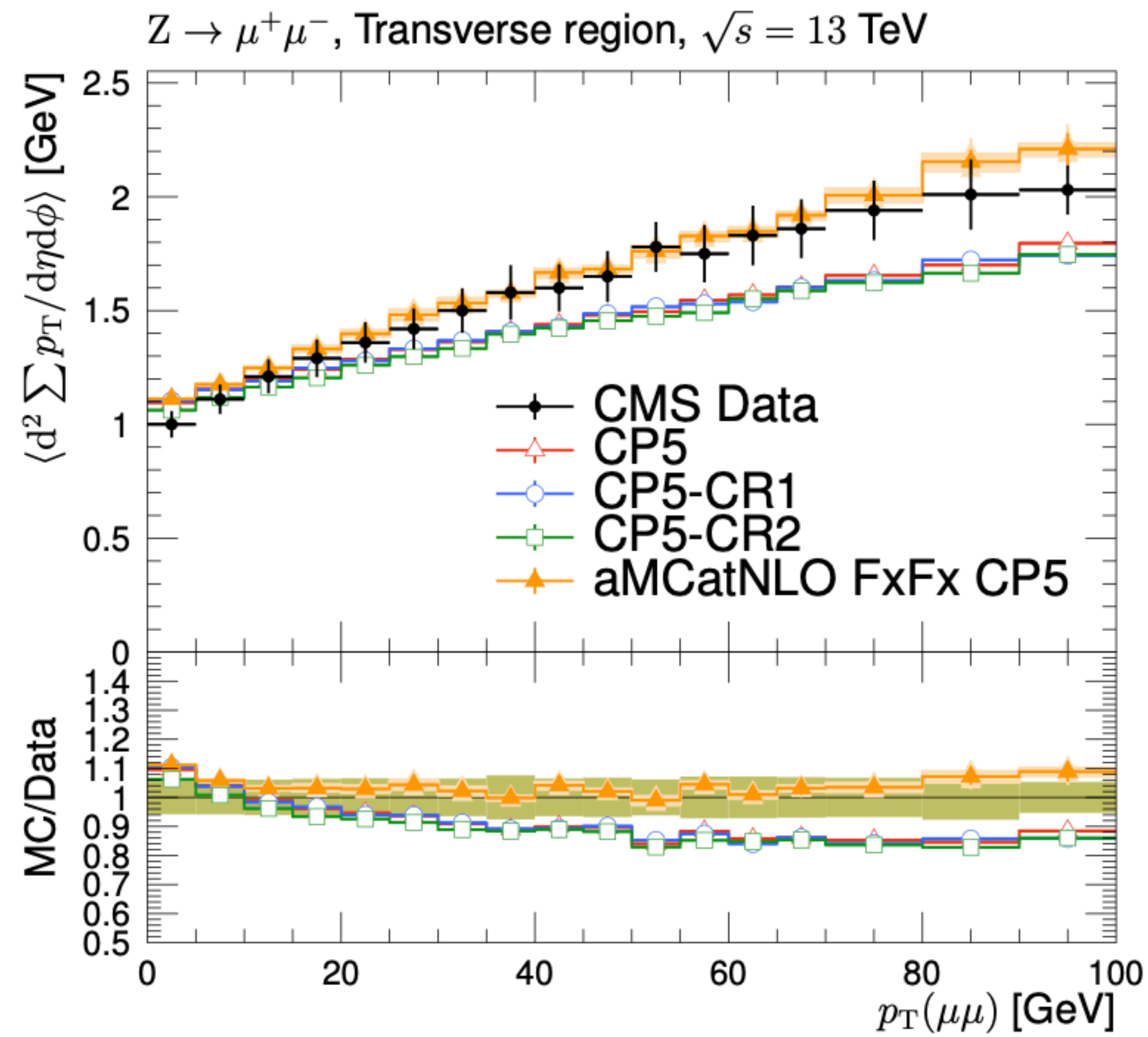
- Average number of charged hadrons with $p_{\mathrm{T}} > 500$ MeV inside jets as a function of jet p_{T}.
- Predictions are comparable.
- CR tunes produce ~5% less charged particles than the CP5 tune.

- $F(z)$ is related to fragmentation function.
- Presented for low $p_{\mathrm{T}}^{\mathrm{jet}} = 25 - 40$ and high $p_{\mathrm{T}}^{\mathrm{jet}} = 400 - 500$ GeV.
- CR tunes describe low-$p_{\mathrm{T}}^{\mathrm{jet}}$ data better than CP5. They also reasonably describe high-$p_{\mathrm{T}}^{\mathrm{jet}}$ data.
- high-$p_{\mathrm{T}}^{\mathrm{jet}}$ data well described by CP5 tune.

z: longitudinal momentum fraction.

Tune Performances: Drell-Yan events (13 TeV)





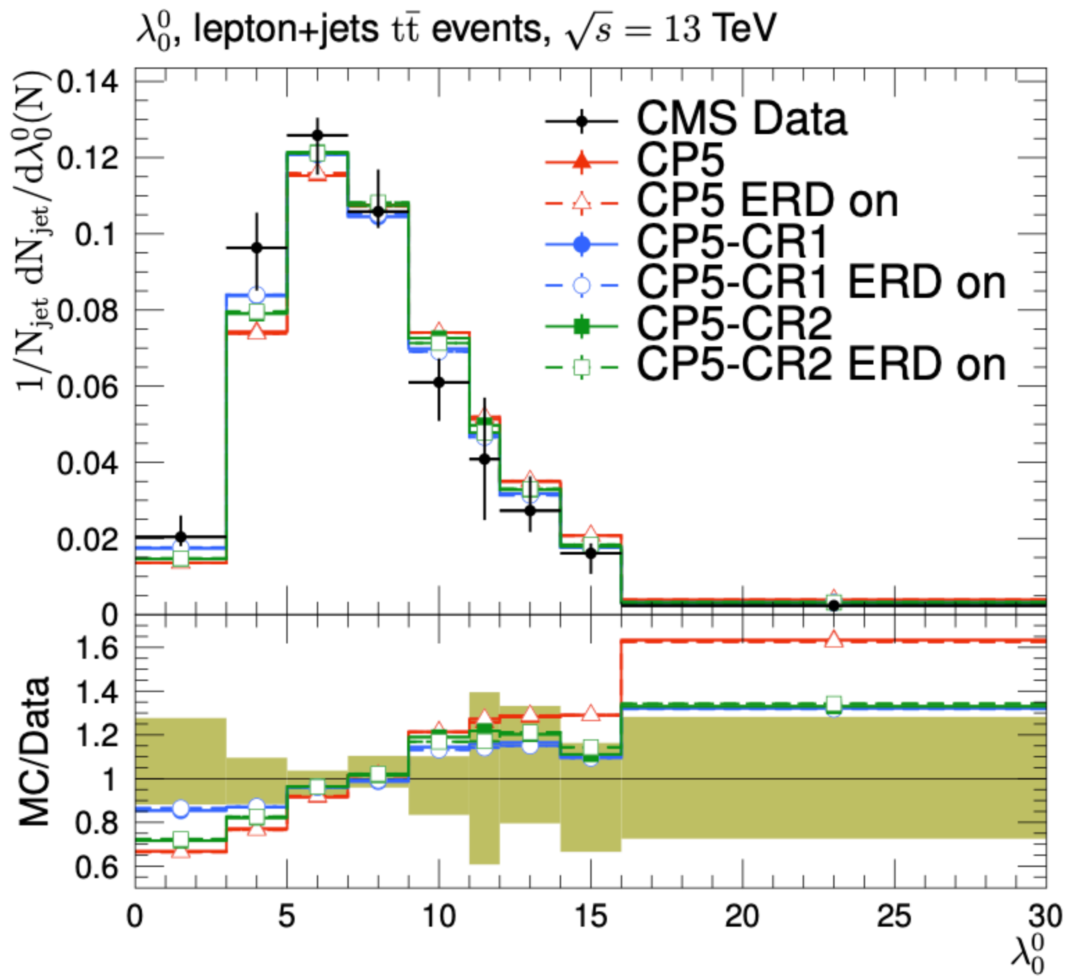
$Z \rightarrow \mu^+\mu^-$ events. N_{ch} and p_{T} flow as a function of Z boson p_{T}.

- CP5 predicts up to 15% more charged particles at low Z boson p_{T}.
  - →But additional effects such as primordial kT expected to play a role.
- Higher order corrections (such as implemented in MG5_aMC@NLO with FxFx) are needed to describe total p_{T} flow.

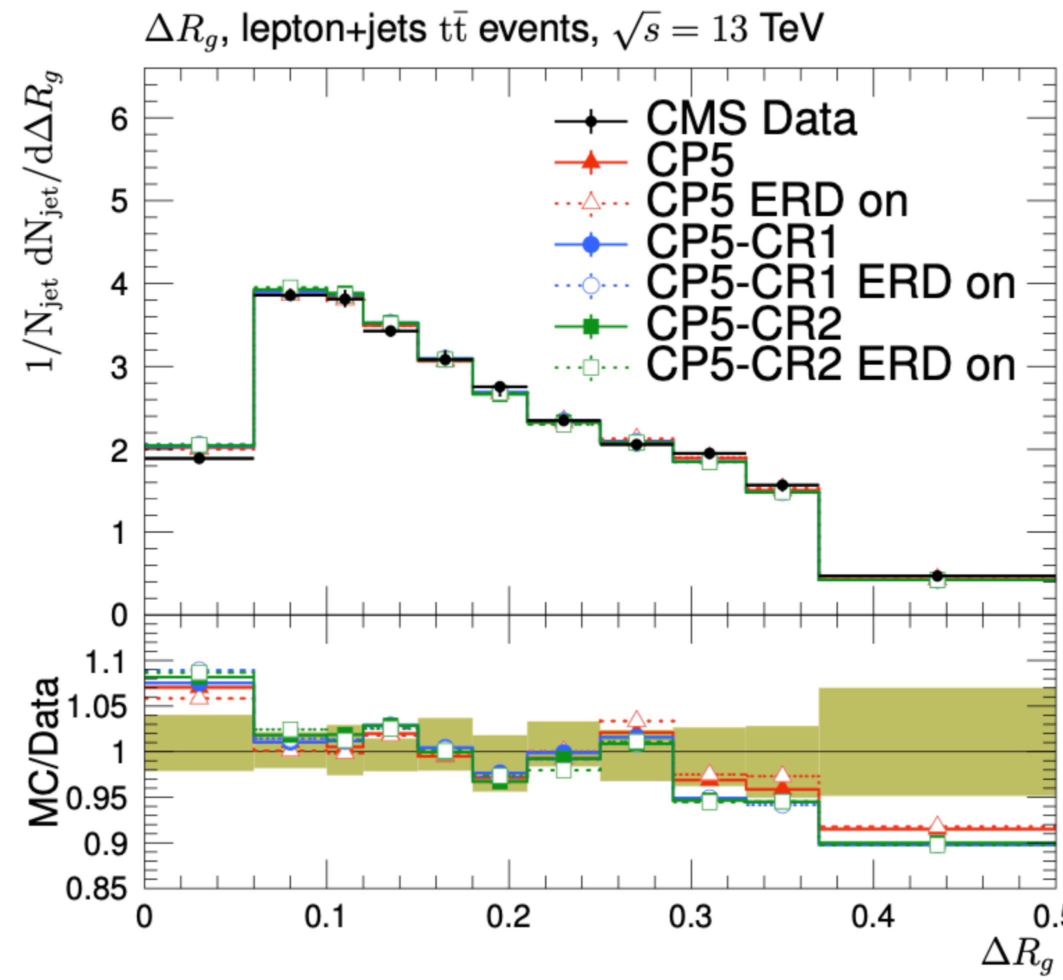
Tune Performances: $t\bar{t}$ jet substructure (13 TeV)

- CR models have no significant effect on UE observables in $t\bar{t}$ events. (CMS-TOP-17-015, EPJC 79 (2019) 123).
- Jet-shape distributions in $t\bar{t}$ events are sensitive CR effects. (CMS-TOP-17-013, Phys. Rev. D 98, 092014 (2018)). Predictions of new tunes tested with these observables.
- $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV, $|\eta| < 2$, $\Delta R(j_1, j_2) > 0.8$. Jet observables are calculated from jet constituents with $p_{\mathrm{T}} > 1$ GeV.

Charged particle multiplicity in the jet



Angle between two groomed jets

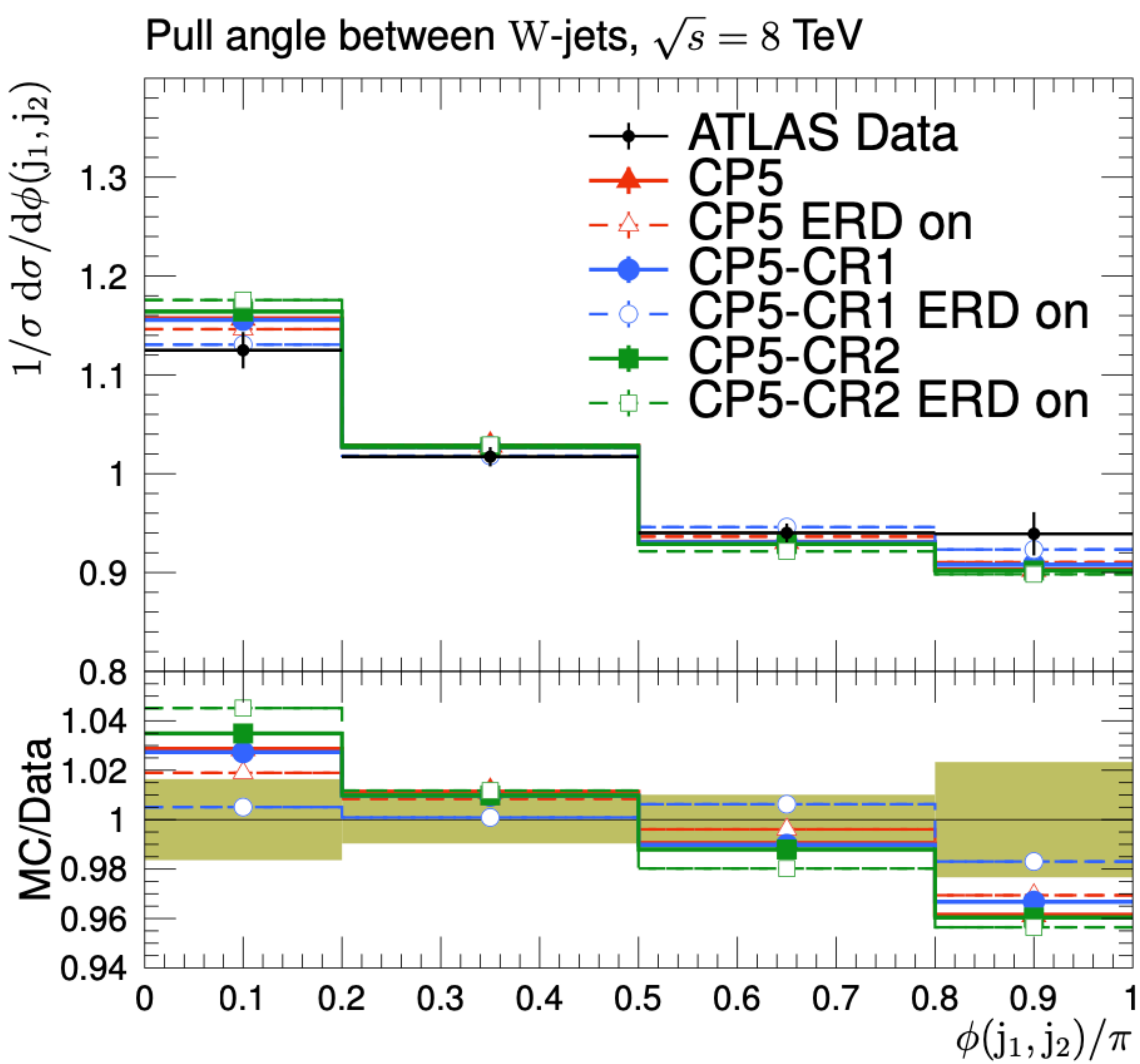


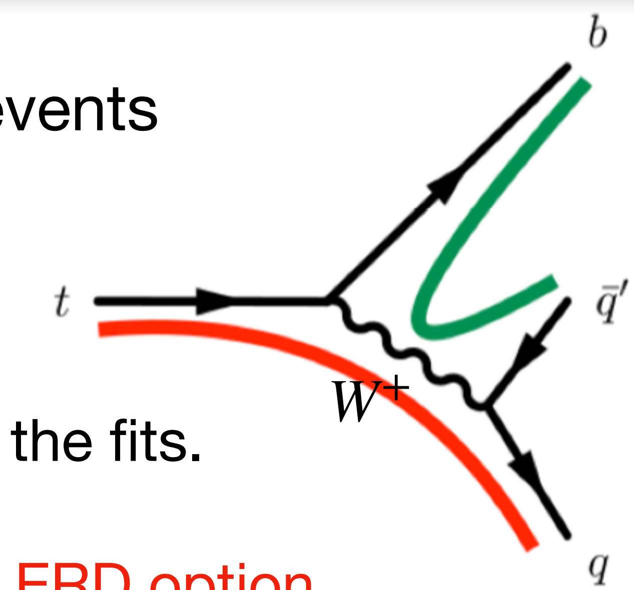
- None of the tunes describe the data well.

groomed jet: a jet with soft and wide-angle radiation removed by a dedicated grooming algorithm.

Tune Performances: $t\bar{t}$ color flow (8 TeV)

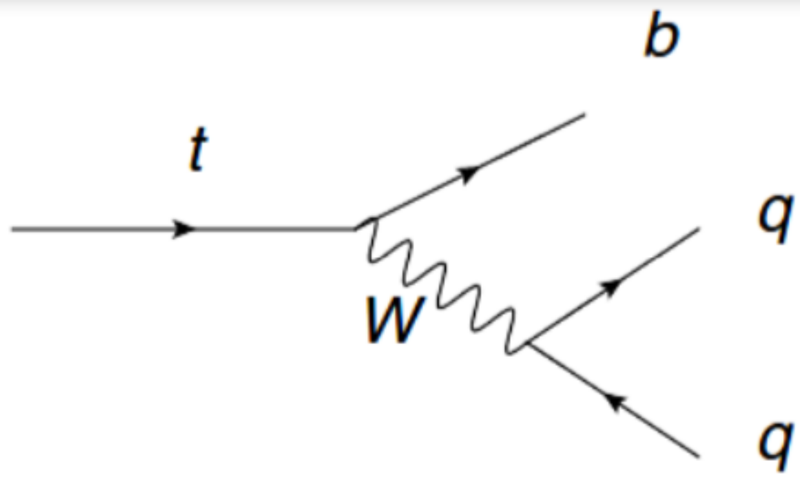
Pull angle between jets originating from the decay of W boson in $t\bar{t}$ events





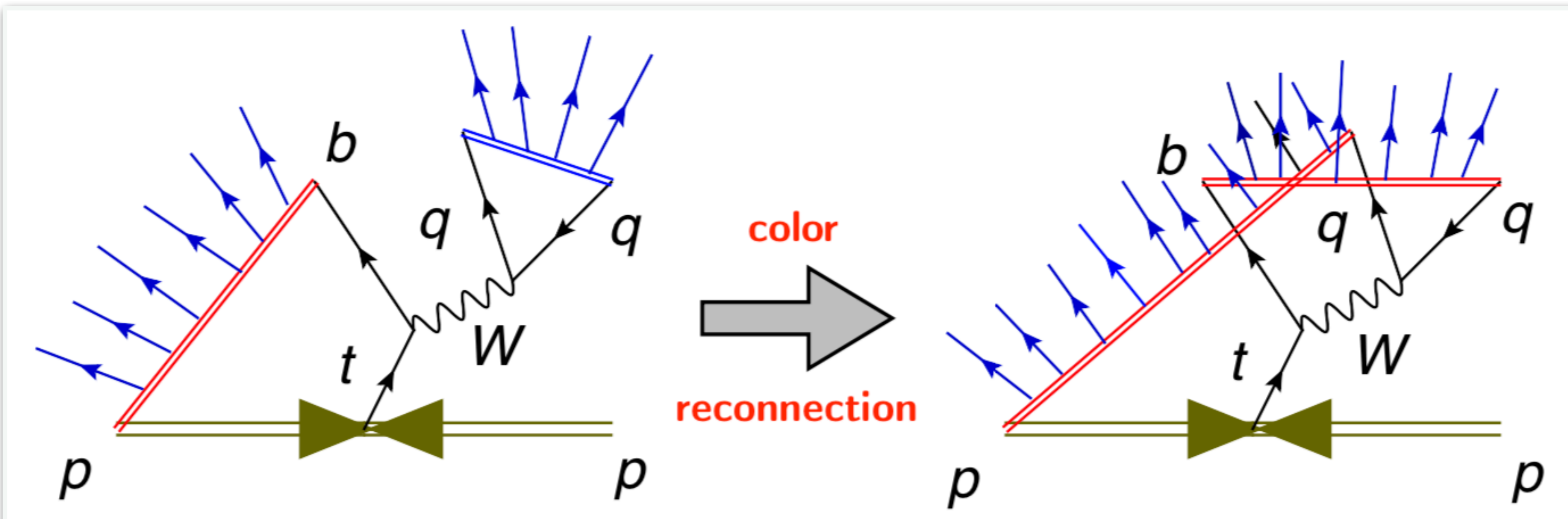
- Not used in the fits.
- Sensitive to ERD option.
- If **ERD off**, decay products of W boson are not included in CR. In this case, tune predictions are very similar.
- If **ERD on**, CR effects modify the pull angle and tune predictions differ a lot.
- CR1 tunes with ERD on provides the best description.
- CR2 tunes prefer ERD off.

Effect of CR on m_{top}



$$t \to bW \to bq\bar{q}$$

$$\widehat{m}^2_{\text{top}} = (p_{\text{b}} + p_{\text{j1}} + p_{\text{j2}})^2$$



Credits: Spyros Argyropoulos
JHEP11 (2014) 043

$$m^2_{\text{top}} \neq (p_{\text{b}} + p_{\text{j1}} + p_{\text{j2}})^2$$

Typical hadronization scale is around 1 fm → But top quark travels ~0.2 fm before it decays.

In PYTHIA8 (Early Resonance Decay (ERD))

- **ERD = off**: top quark can colour reconnect to other partons.
- **ERD = on**: the decay products of the top quark can colour reconnect to other partons.

Estimating the CR uncertainty on m_{top}

- The top quark mass has been measured with high precision using the 7, 8, and 13 TeV data at the LHC.
- A complete analysis of the systematic uncertainties in the measurement is crucial.
- One of the most dominant systematic uncertainty is due to CR.

LHC Run 1: Uncertainty on top mass due to CR effects is usually studied by simply taking the difference of

$$\Delta m_{\text{top}} = \hat{m}_{\text{top}}^{\text{default}} - \hat{m}_{\text{top}}^{\text{no-CR}}$$

The Problem: no-CR is unphysical and does not describe the underlying soft physics of pp collisions.

LHC Run 2: Estimate uncertainty by comparing different CR models and settings.

+

Compare effects of ERD on/off option in CR models

Need(ed) a model tuned to existing data

Top quark mass predicted by different tunes

- Most precise measurement by CMS experiment combining the data at 7 and 8 TeV:

[CMS-TOP-17-007, PhysRevD 93.072004]

$$m_t = 172.44 \pm 0.13\,(\text{stat+JSF}) \pm 0.47\,(\text{syst})\,\text{GeV},$$

Used in CMS-TOP-17-007 to estimate uncertainty on m_{top}

$$\Delta m_t - 0.5\Delta m_W$$

Tune	m_t [GeV]	Δm_t [GeV]	m_W [GeV]	Δm_W [GeV]	Δm_t^{hyb} [GeV]
CP5	171.93 ± 0.02	—	79.76 ± 0.02	—	—
CP5 ERD	172.18 ± 0.03	0.25	80.15 ± 0.02	0.40	0.05
CP5-CR1	171.97 ± 0.02	0.04	79.74 ± 0.02	−0.02	0.05
CP5-CR1 ERD	172.01 ± 0.03	0.08	79.98 ± 0.02	0.23	−0.04
CP5-CR2	171.91 ± 0.02	−0.02	79.85 ± 0.02	0.10	−0.07
CP5-CR2 ERD	172.32 ± 0.03	0.39	79.90 ± 0.02	0.14	0.32

largest deviation from the predictions of CP5

- Top/W mass values obtained by fitting a Gaussian function within an 8 GeV mass window around the corresponding mass peak.
- 0.32 GeV is similar to the shift (0.31 GeV) found in CMS-TOP-17-007 using CUETP8M2T4. *CP5 does not improve or degrade the precision of the top quark mass measurements.*

Summary

- New set of parameters for two of the CR models, QCD-inspired and gluon-move, implemented in the PYTHIA8 are obtained, based on the default CMS PYTHIA8 tune CP5.
- Measurements sensitive to UE contributions performed at 1.96, 7 and 13 TeV are used to constrain the parameters for the CR and for the MPI simultaneously.
- Various measurements at 1.96, 7, 8 and 13 TeV, as well as data from e^+e^- colliders are used to evaluate the performance of the new tunes.
    - New CR models alone do not improve the description of strange particle production as a function of rapidity for Λ baryons. K^0_S versus rapidity described well by all tunes (CP5, CP5-CR1, and CP5-CR2).
    - None of the tunes describe jet shapes in $t\bar{t}$ events well. Besides, all tunes give similar predictions.
- Including jet pull angle and other jet substructure observables in $t\bar{t}$ could beneficial for further tuning studies.
- A study on the uncertainty of the top quark mass measurement due to CR effects is also presented.
    - **CR is still the dominant source of uncertainty in m_{top} measurements.**

Backup

CUETP8M2T4 based CR tunes

Parameters	CUETP8M2T4	QCD inspired	gluon move
MultipartonInteractions:**pT0Ref**	2.20	2.17	2.30
MultipartonInteractions:**expPow**	1.60	1.31	1.35
MultipartonInteractions:**ecmRef**	7000	7000*	7000*
MultipartonInteractions:**ecmPow**	0.25	0.25*	0.25*
ColourReconnection:**range**	6.59	-	-
ColourReconnection:**junctionCorrection**	-	0.12 (1.20)	-
ColourReconnection:**timeDilationPar**	-	15.9 (0.18)	-
ColourReconnection:**m0**	-	1.2 (0.3)	-
ColourReconnection:**m2lambda**	-	-	1.9 (1.0)
ColourReconnection:**fracGluon**	-	-	1.0* (1.0)
ColourReconnection:**dLambdaCut**	-	-	0.0* (0.0)
PDF set	NNPDF30_LO [JHEP 04 (2015)]	NNPDF30_LO	NNPDF30_LO
SpaceShower:alphaSvalue	0.1108*	0.1108*	0.1108*
Goodness of fit/dof	1.89 [CMS-PAS-TOP-16-021]	1.06	1.69
* = value kept fixed in the fit			

- A study based on early UE tune of Run 2. Tunes have been used used in Ref.[1] to calculate the uncertainty on top mass due to the CR modeling.
- NNPDF3.0 LO PDF set.
- A simpler impact parameter profile for the incoming hadron beams is used, i.e., double Gaussion matter is not used.
- In gluon-move model, only m2Lambda parameter was tuned.

[1] CMS Collaboration, "Measurement of the top quark mass with lepton+jets final states using pp collisions at s = 13 TeV", *Eur. Phys. J. C* **78** (2018) 891.