

Monte Carlo modeling and tuning in CMS

Gerrit Van Onsem
on behalf of the CMS Collaboration

40th International Conference on High Energy Physics
Prague, Czech Republic (virtual)

30 July 2020

Simulating a hadron-hadron collision

Monte Carlo (MC) event generators simulate

Hard scattering matrix element (ME)

Parton showers (PS) / radiation

Hadronization / hadron decays

Multiple-parton interactions (MPI)

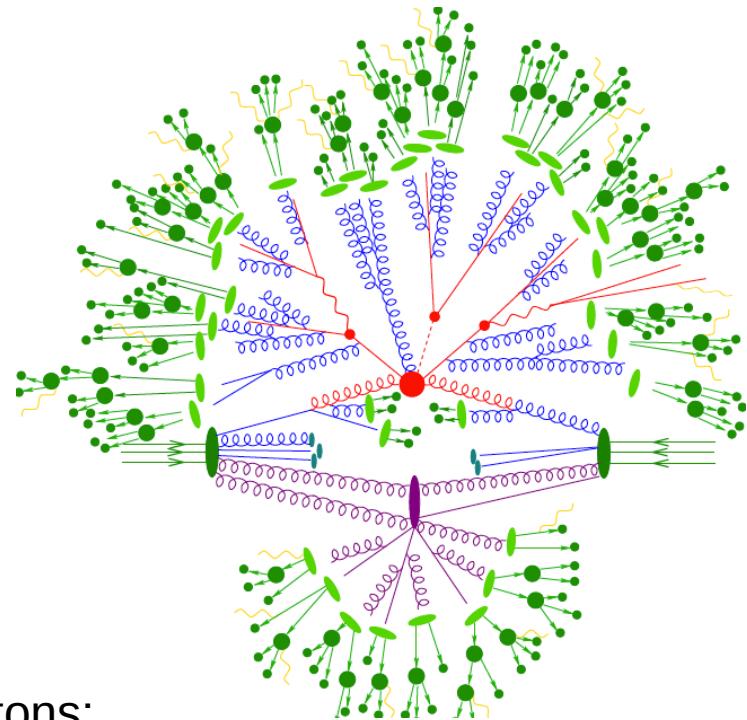
Beam remnants



“Underlying event” (UE)

involving color exchange between partons;

color fields may be rearranged → color reconnection (CR)



Commonly used **MC generators** in CMS include

Powheg, MadGraph5_aMC@NLO for fixed-order ME calculations

Sherpa multipurpose

Pythia, Herwig multipurpose, but often used to interface parton shower and UE simulation to ME

Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

Pythia 8.2 [GEN-17-001](#)

Parameter description	Range considered
MPI threshold [GeV], $p_{\text{T}0\text{Ref}}$, at $\sqrt{s} = \sqrt{s_0}$	1.0–3.0
Exponent of \sqrt{s} dependence, ϵ	0.0–0.3
Matter fraction contained in the core	0.1–0.95
Radius of the core	0.1–0.8
Range of color reconnection probability	1.0–9.0

Herwig 7.1 [GEN-19-001](#)

Parameter	Range	
$p_{\perp,0}^{\min}$ (GeV)	1.0–5.0	
b (GeV)	0.1–0.5	
μ^2 (GeV $^{-2}$)	0.5–2.7	
p_{reco}	0.05–0.90	

Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

- p_T threshold parameter (p_T^0) to govern transition between soft and hard interactions; has \sqrt{s} energy dependence
lower threshold → more MPI, i.e. more UE activity

Pythia 8.2 [GEN-17-001](#)

Parameter description	Range considered
MPI threshold [GeV], $p_{T0\text{Ref}}$, at $\sqrt{s} = \sqrt{s_0}$	1.0–3.0
Exponent of \sqrt{s} dependence, ϵ	0.0–0.3
Matter fraction contained in the core	0.1–0.95
Radius of the core	0.1–0.8
Range of color reconnection probability	1.0–9.0

Herwig 7.1 [GEN-19-001](#)

Parameter	Range
$p_{\perp,0}^{\min}$ (GeV)	1.0–5.0
b (GeV)	0.1–0.5
μ^2 (GeV $^{-2}$)	0.5–2.7
p_{reco}	0.05–0.90



Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

- p_T threshold parameter (p_T^0) to govern transition between soft and hard interactions; **has \sqrt{s} energy dependence**
lower threshold → more MPI, i.e. more UE activity

Pythia 8.2 [GEN-17-001](#)

Parameter description	Range considered
MPI threshold [GeV], pT0Ref, at $\sqrt{s} = \sqrt{s_0}$	1.0–3.0
Exponent of \sqrt{s} dependence, ϵ	0.0–0.3
Matter fraction contained in the core	0.1–0.95
Radius of the core	0.1–0.8
Range of color reconnection probability	1.0–9.0

Herwig 7.1 [GEN-19-001](#)

Parameter	Range
$p_{\perp,0}^{\min}$ (GeV)	1.0–5.0
b (GeV)	0.1–0.5
μ^2 (GeV $^{-2}$)	0.5–2.7
p_{reco}	0.05–0.90



Exponent

Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

- p_T threshold parameter (p_T^0) to govern transition between soft and hard interactions; has \sqrt{s} energy dependence
lower threshold → *more MPI, i.e. more UE activity*
- Overlap distribution between two colliding protons
larger / denser overlap → *more MPI*

Pythia 8.2 [GEN-17-001](#)

Parameter description	Range considered
MPI threshold [GeV], pT0Ref, at $\sqrt{s} = \sqrt{s_0}$	1.0–3.0
Exponent of \sqrt{s} dependence, ϵ	0.0–0.3
Matter fraction contained in the core	0.1–0.95
Radius of the core	0.1–0.8
Range of color reconnection probability	1.0–9.0

Herwig 7.1 [GEN-19-001](#)

Parameter	Range
$p_{\perp,0}^{\min}$ (GeV)	1.0–5.0
b (GeV)	0.1–0.5
μ^2 (GeV $^{-2}$)	0.5–2.7
p_{reco}	0.05–0.90



Inverse proton radius squared

Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

- p_T threshold parameter (p_T^0) to govern transition between soft and hard interactions; has \sqrt{s} energy dependence
lower threshold → *more MPI, i.e. more UE activity*
- Overlap distribution between two colliding protons
larger / denser overlap → *more MPI*
- Color reconnection probability
large value → *tends to reduce final particle multiplicities*

Pythia 8.2 [GEN-17-001](#)

Parameter description	Range considered
MPI threshold [GeV], pT0Ref, at $\sqrt{s} = \sqrt{s_0}$	1.0–3.0
Exponent of \sqrt{s} dependence, ϵ	0.0–0.3
Matter fraction contained in the core	0.1–0.95
Radius of the core	0.1–0.8
Range of color reconnection probability	1.0–9.0

Herwig 7.1 [GEN-19-001](#)

Parameter	Range
$p_{\perp,0}^{\min}$ (GeV)	1.0–5.0
b (GeV)	0.1–0.5
μ^2 (GeV $^{-2}$)	0.5–2.7
p_{reco}	0.05–0.90



Tuning parton shower generators

Underlying event, parton shower development, hadronization (\sim lower scales)
not always calculable in perturbative QCD

→ modeling governed by **phenomenological parameters** that can be “tuned”

- p_T threshold parameter (p_T^0) to govern transition between soft and hard interactions; has \sqrt{s} energy dependence
lower threshold → *more MPI, i.e. more UE activity*
- Overlap distribution between two colliding protons
larger / denser overlap → *more MPI*
- Color reconnection probability
large value → *tends to reduce final particle multiplicities*

Parton distribution functions (PDF), strong coupling (α_s) affect predictions too

Tuning performed by **fitting predictions to data** using Rivet + Professor

$$\chi^2(p) = \sum_{\mathcal{O}} w_{\mathcal{O}} \sum_{i \in \mathcal{O}} \frac{(f^i(p) - \mathcal{R}_i)^2}{\Delta_i^2}$$

p = tuning parameters

f^i = predicted bin content from parametrization

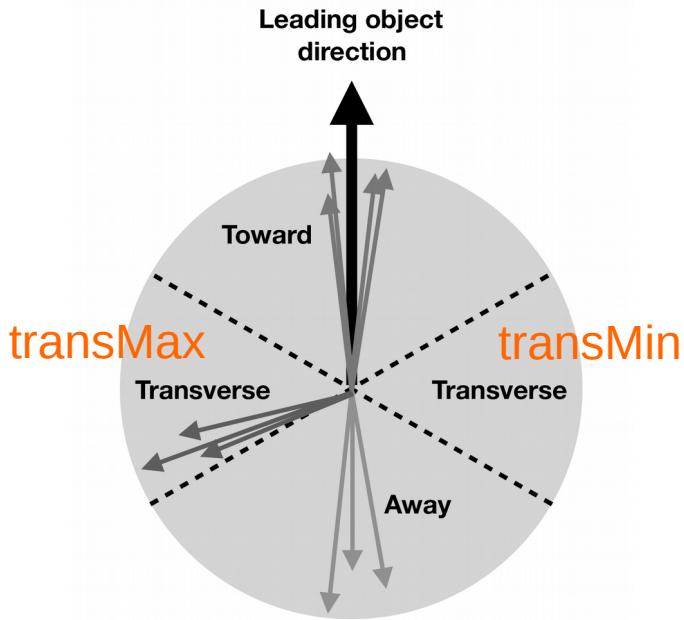
R_i = measured bin content

\mathcal{O} = binned observable

Observables sensitive to UE

Tunes extracted from relevant observables in unfolded **minimum-bias (MB) data**

= inelastic events collected with loose selection

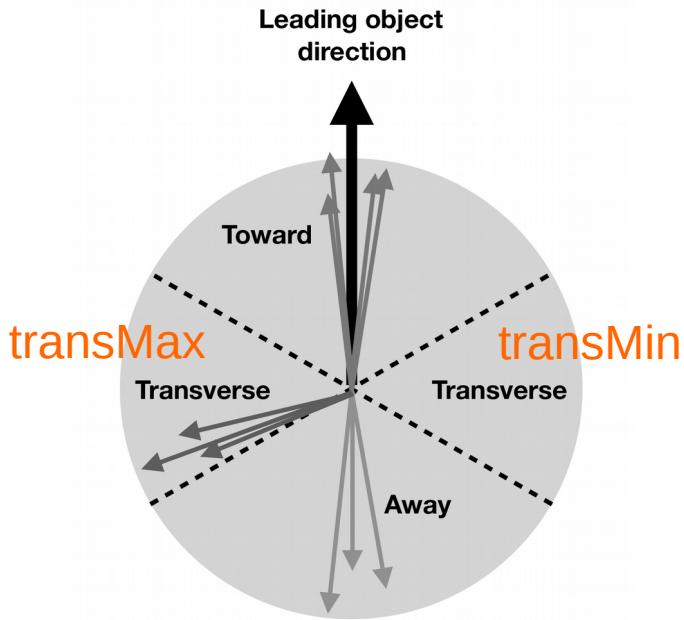


- **Charged-particle multiplicity (N_{ch}) and scalar p_{T} sum ($p_{\text{T}}^{\text{sum}}$) densities as function of $p_{\text{T}} > 3 \text{ GeV}$ of leading object (= track or charged-particle jet), in transverse regions**
CMS @ 7 + 13 TeV (+ 0.9 TeV for Herwig)
(+ CDF @ 1.96 TeV for Pythia)
- **N_{ch} as function of charged hadron η**
CMS @ 13 TeV

Observables sensitive to UE

Tunes extracted from relevant observables in unfolded **minimum-bias (MB) data**

= inelastic events collected with loose selection



- **Charged-particle multiplicity (N_{ch}) and scalar p_{T} sum ($p_{\text{T}}^{\text{sum}}$) densities as function of $p_{\text{T}} > 3 \text{ GeV}$ of leading object (= track or charged-particle jet), in transverse regions**
CMS @ 7 + 13 TeV (+ 0.9 TeV for Herwig)
(+ CDF @ 1.96 TeV for Pythia)
- **N_{ch} as function of charged hadron η**
CMS @ 13 TeV

Components of hard scattering such as hadronization and initial-state (ISR) and final-state radiation (FSR) may affect UE-sensitive observables

→ transMax more sensitive to ISR & FSR; transMin to MPI & beam remnants

Predictions obtained by **simulating non-diffractive + diffractive inelastic events** for each choice of tune parameters

Simulation with Pythia 8 includes central diffraction

Results of CMS Pythia 8 tunes

Different assumptions for **order of PDF set, and α_s value + running**

LO-PDF tunes

PYTHIA8 parameter	CP1 NNPDF3.1 LO	CP2 NNPDF3.1 LO
PDF Set		
$\alpha_s(m_Z)$	0.130	0.130
SpaceShower:rapidityOrder	off	off
MultipartonInteractions:EcmRef [GeV]	7000	7000
$\alpha_s^{\text{ISR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_s^{\text{FSR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_s^{\text{MPI}}(m_Z)$ value/order	0.130/LO	0.130/LO
$\alpha_s^{\text{ME}}(m_Z)$ value/order	0.130/LO	0.130/LO
MultipartonInteractions:pT0Ref [GeV]	2.4	2.3
MultipartonInteractions:ecmPow	0.15	0.14
MultipartonInteractions:coreRadius	0.54	0.38
MultipartonInteractions:coreFraction	0.68	0.33
ColorReconnection:range	2.63	2.32
χ^2/dof	0.89	0.54

- Higher-order PDF used in NLO ME may motivate usage of higher-order PDF in PS / UE
- Different assumptions allow to **check consistency of matching order of PDF** used in ME and in tunes

(N)NLO-PDF tunes

PYTHIA8 parameter	CP3 NNPDF3.1 NLO	CP4 NNPDF3.1 NNLO	CP5 NNPDF3.1 NNLO
PDF Set			
$\alpha_s(m_Z)$	0.118	0.118	0.118
SpaceShower:rapidityOrder	off	off	on
MultipartonInteractions:EcmRef [GeV]	7000	7000	7000
$\alpha_s^{\text{ISR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{FSR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{MPI}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{ME}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
MultipartonInteractions:pT0Ref [GeV]	1.52	1.48	1.41
MultipartonInteractions:ecmPow	0.02	0.02	0.03
MultipartonInteractions:coreRadius	0.54	0.60	0.76
MultipartonInteractions:coreFraction	0.39	0.30	0.63
ColorReconnection:range	4.73	5.61	5.18
χ^2/dof	0.76	0.80	1.04

Rapidity ordering of ISR reduces phase space for parton emissions

Results of CMS Pythia 8 tunes

Different assumptions for **order of PDF set**, and α_s value + running

LO-PDF tunes

PYTHIA8 parameter	CP1 NNPDF3.1 LO	CP2 NNPDF3.1 LO
PDF Set		
$\alpha_s(m_Z)$	0.130	0.130
SpaceShower:rapidityOrder	off	off
MultipartonInteractions:EcmRef [GeV]	7000	7000
$\alpha_s^{\text{ISR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_s^{\text{FSR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_s^{\text{MPI}}(m_Z)$ value/order	0.130/LO	0.130/LO
$\alpha_s^{\text{ME}}(m_Z)$ value/order	0.130/LO	0.130/LO
MultipartonInteractions:pT0Ref [GeV]	2.4	2.3
MultipartonInteractions:ecmPow	0.15	0.14
MultipartonInteractions:coreRadius	0.54	0.38
MultipartonInteractions:coreFraction	0.68	0.33
ColorReconnection:range	2.63	2.32
χ^2/dof	0.89	0.54

Significant differences between LO- and (N)NLO-based tunes in value of p_T^0 and energy dependence, and amount of CR:

- Different shapes of gluon densities at small x
- α_s running

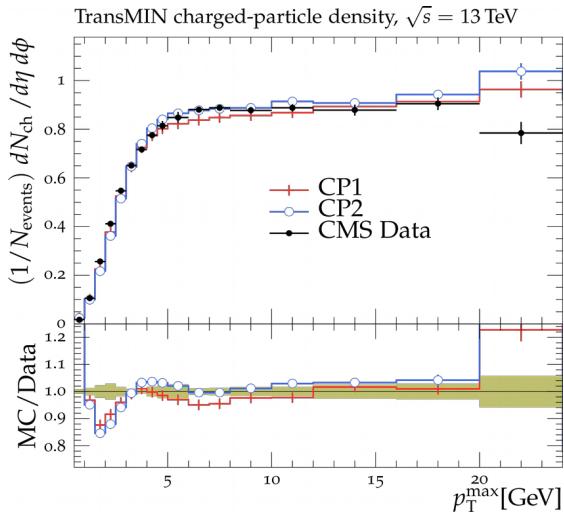
(N)NLO-PDF tunes

PYTHIA8 parameter	CP3 NNPDF3.1 NLO	CP4 NNPDF3.1 NNLO	CP5 NNPDF3.1 NNLO
PDF Set			
$\alpha_s(m_Z)$	0.118	0.118	0.118
SpaceShower:rapidityOrder	off	off	on
MultipartonInteractions:EcmRef [GeV]	7000	7000	7000
$\alpha_s^{\text{ISR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{FSR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{MPI}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_s^{\text{ME}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
MultipartonInteractions:pT0Ref [GeV]	1.52	1.48	1.41
MultipartonInteractions:ecmPow	0.02	0.02	0.03
MultipartonInteractions:coreRadius	0.54	0.60	0.76
MultipartonInteractions:coreFraction	0.39	0.30	0.63
ColorReconnection:range	4.73	5.61	5.18
χ^2/dof	0.76	0.80	1.04

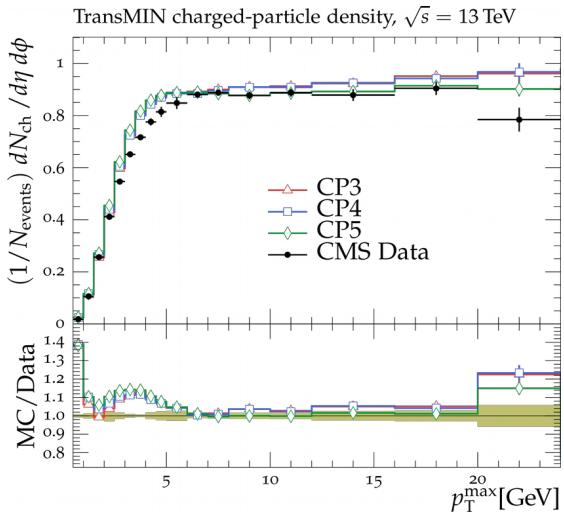
Rapidity ordering of ISR reduces phase space for parton emissions

Example observables used in Pythia 8 tuning

LO-PDF tunes



(N)NLO-PDF tunes

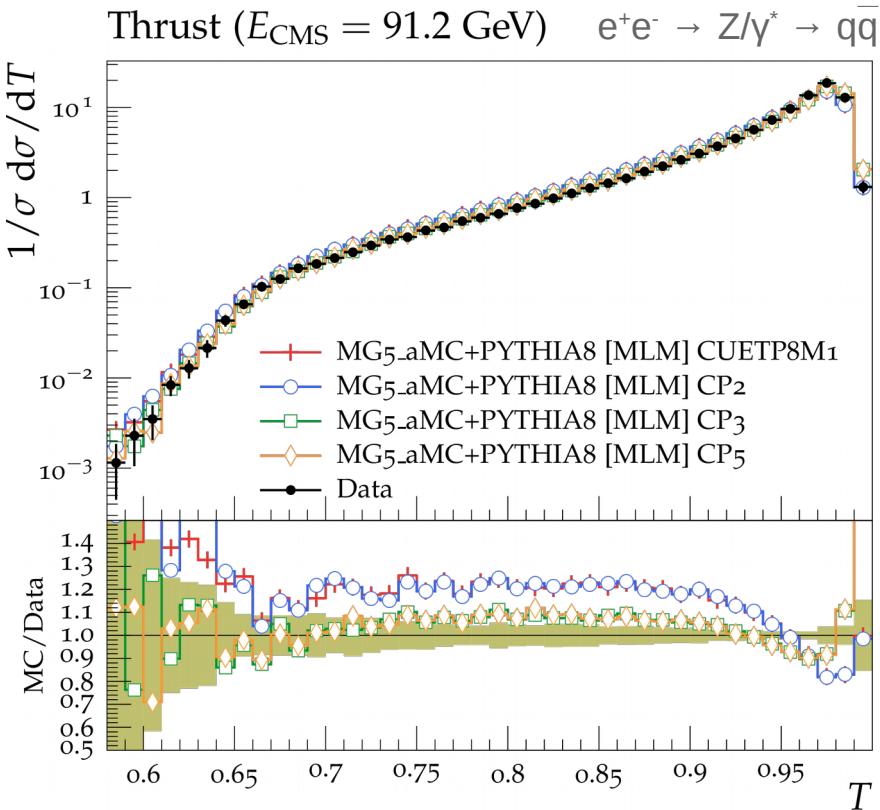


- Rising part of spectrum: diffraction processes become more important
 - not always well described; $p_T^{\max} < 3 \text{ GeV}$ not included in tuning
- No MB or UE observable where level of agreement between data and prediction from different tunes is significantly different

Example validations of different Pythia 8 tunes

Ideally, UE tune is **universal** and describes wide range of processes & observables

LEP event shape observables

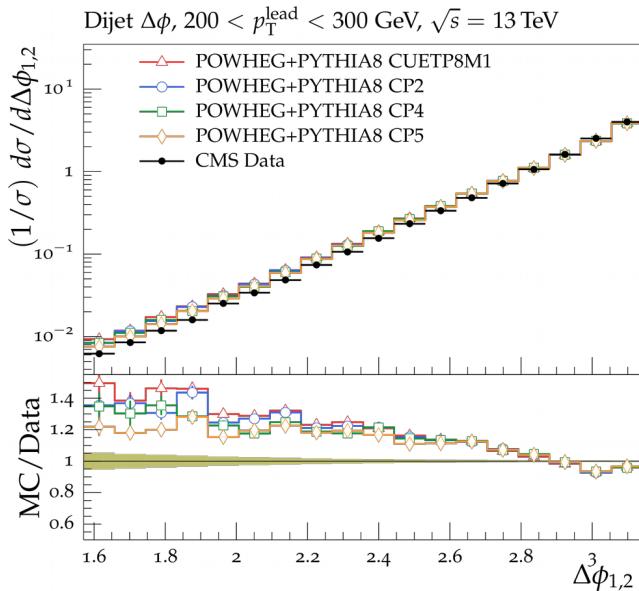


- Predictions from MG₅.aMC@NLO + Pythia 8 with up to 4 partons at ME
- Leptonic initial state → **observables particularly sensitive to α_s^{FSR}**
- Thrust (T): dijet event $T \sim 1$, isotropic event $T \sim 0.5$
- **Tunes assuming lower α_s^{FSR} describe data better** (less FSR → less isotropic)

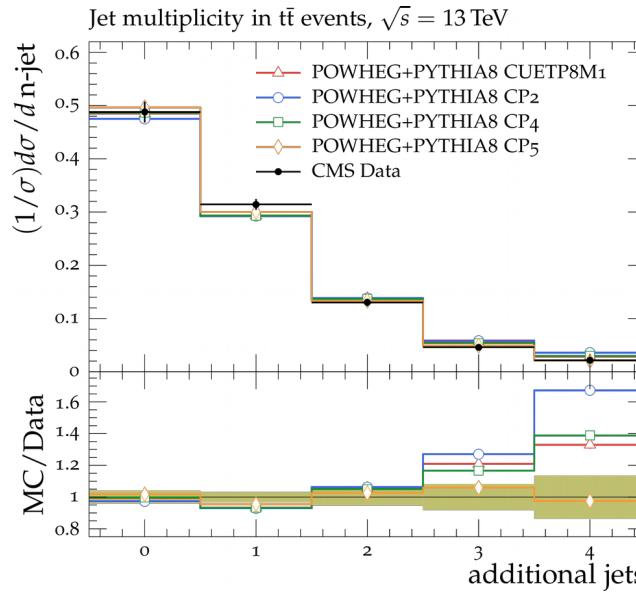
Example validations of different Pythia 8 tunes

Ideally, UE tune is **universal** and describes wide range of processes & observables

Multijet final states



Top quark pair production



- **NLO dijet ME in Powheg merged with Pythia 8**
- **Tunes based on NLO α_s running better** ($\text{lower } \alpha_s^{\text{FSR}} \rightarrow \text{less jet decorrelation}$) than tunes with LO running
- Interfacing **NLO Powheg or MG5_aMC@NLO ME with Pythia 8**
- **$\alpha_s = 0.118 + \text{rapidity ordering}$ of ISR (= CP5 tune) favored for Powheg + Pythia 8**

Results of CMS Herwig 7 tunes

Different assumptions for **order of PDF set in UE, and α_s value + running**

		SoftTune	CH1	CH2	CH3
	$\alpha_s(m_Z)$	0.1262	0.118	0.118	0.118
PS	PDF set $\alpha_s^{\text{PDF}}(m_Z)$	MMHT2014 LO 0.135	NNPDF3.1 NNLO 0.118	NNPDF3.1 NNLO 0.118	NNPDF3.1 NNLO 0.118
MPI & remnants	PDF set $\alpha_s^{\text{PDF}}(m_Z)$	MMHT2014 LO 0.135	NNPDF3.1 NNLO 0.118	NNPDF3.1 LO 0.118	NNPDF3.1 LO 0.130
$p_{\perp,0}^{\min}$ (GeV)		3.502	2.322	3.138	3.040
b (GeV)		0.416	0.157	0.120	0.136
μ^2 (GeV $^{-2}$)		1.402	1.532	1.174	1.284
p_{reco}		0.5	0.400	0.479	0.471
χ^2/N_{dof}		12.8	4.15	1.54	1.71

NNLO PDF for parton shower, consistent with PDF in ME calculations

SoftTune: tune performed by Herwig 7 authors using ATLAS data

Results of CMS Herwig 7 tunes

Different assumptions for **order of PDF set in UE, and α_s value + running**

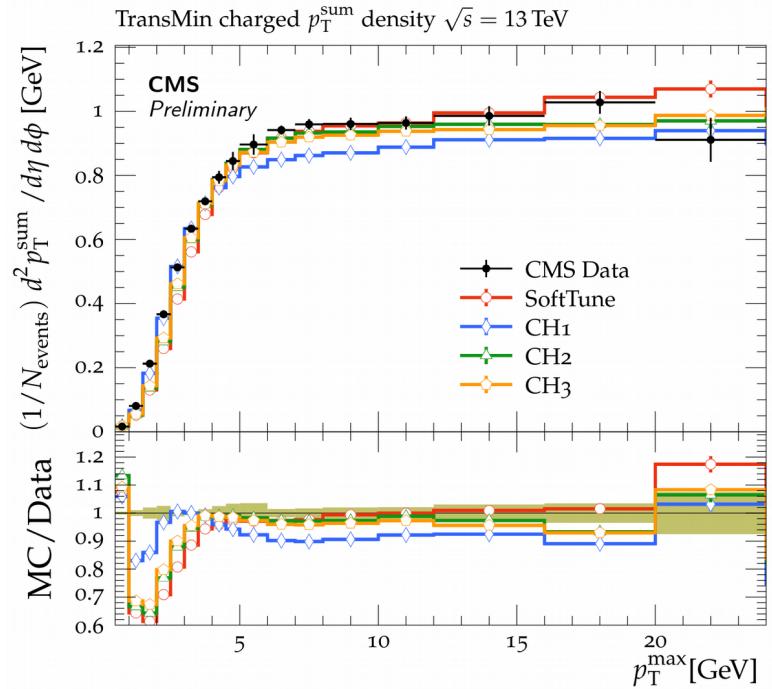
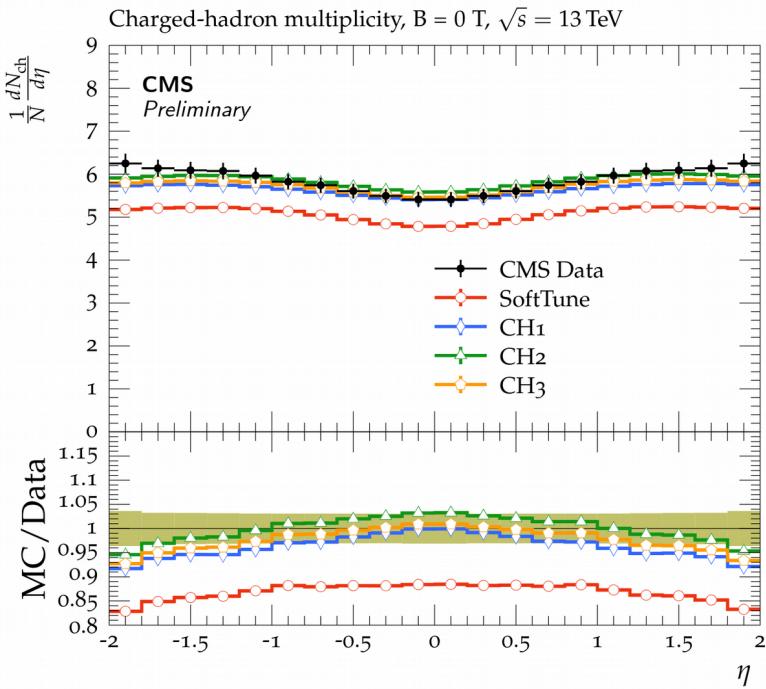
		SoftTune	CH1	CH2	CH3
	$\alpha_s(m_Z)$	0.1262	0.118	0.118	0.118
PS	PDF set $\alpha_s^{\text{PDF}}(m_Z)$	MMHT2014 LO 0.135	NNPDF3.1 NNLO 0.118	NNPDF3.1 NNLO 0.118	NNPDF3.1 NNLO 0.118
MPI & remnants	PDF set $\alpha_s^{\text{PDF}}(m_Z)$	MMHT2014 LO 0.135	NNPDF3.1 NNLO 0.118	NNPDF3.1 LO 0.118	NNPDF3.1 LO 0.130
$p_{\perp,0}^{\min}$ (GeV)	3.502	2.322	3.138	3.040	
b (GeV)	0.416	0.157	0.120	0.136	
μ^2 (GeV $^{-2}$)	1.402	1.532	1.174	1.284	
p_{reco}	0.5	0.400	0.479	0.471	
χ^2/N_{dof}	12.8	4.15	1.54	1.71	

NNLO PDF for parton shower, consistent with PDF in ME calculations

SoftTune: tune performed by Herwig 7 authors using ATLAS data

- Lower p_T^0 and b compared to SoftTune
→ **increased amount of MPI in CH tunes**
- **LO PDF for UE preferred over NNLO PDF**
- **Choice of α_s in PDF less important,**
but $\alpha_s = 0.130$ typically associated with LO PDF

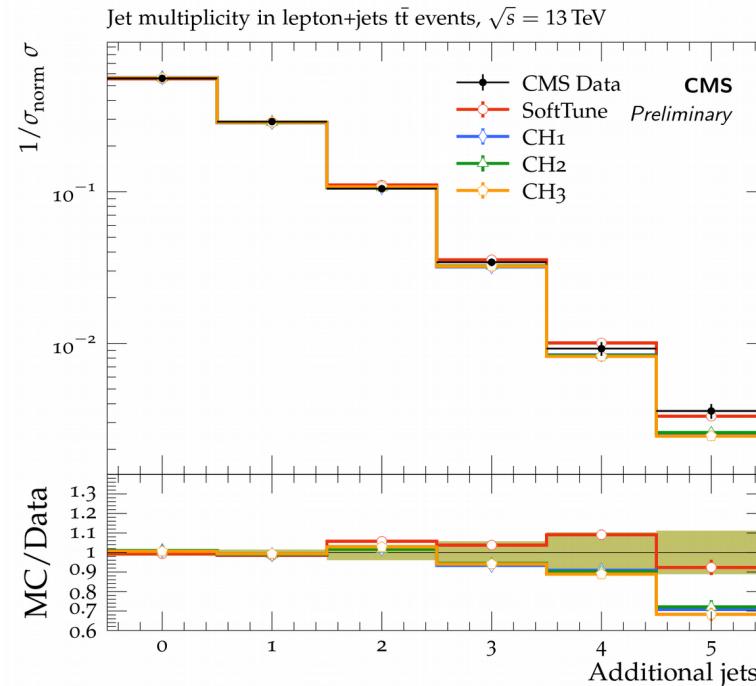
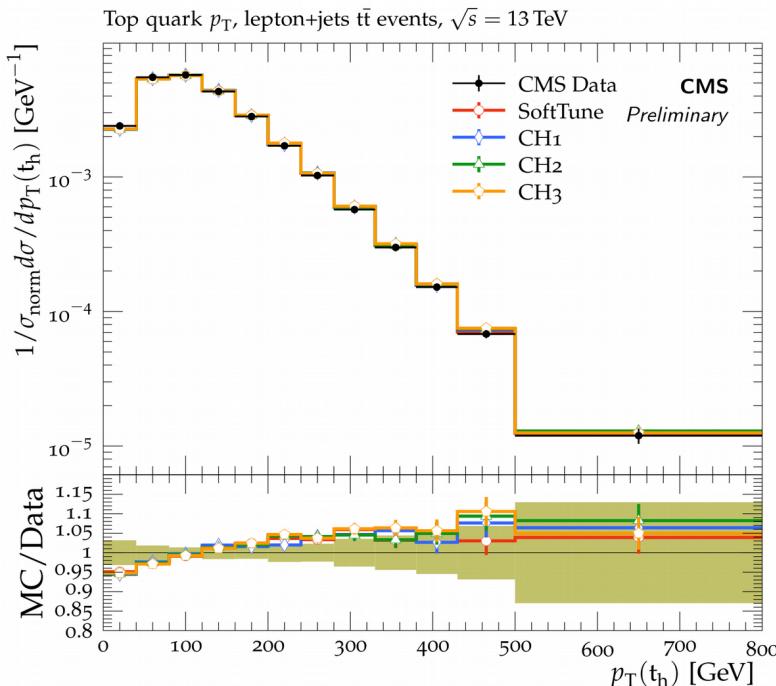
Example observables used in Herwig 7 tuning



- **CH tunes increase amount of MPI w.r.t SoftTune**
- **CH tunes show typically good agreement in plateau of N_{ch} or p_T^{sum} density spectrum**
- **Rising part of spectrum (diffraction processes become more important) not included in tuning procedure**

Example validations of different Herwig 7 tunes

Top quark pair production

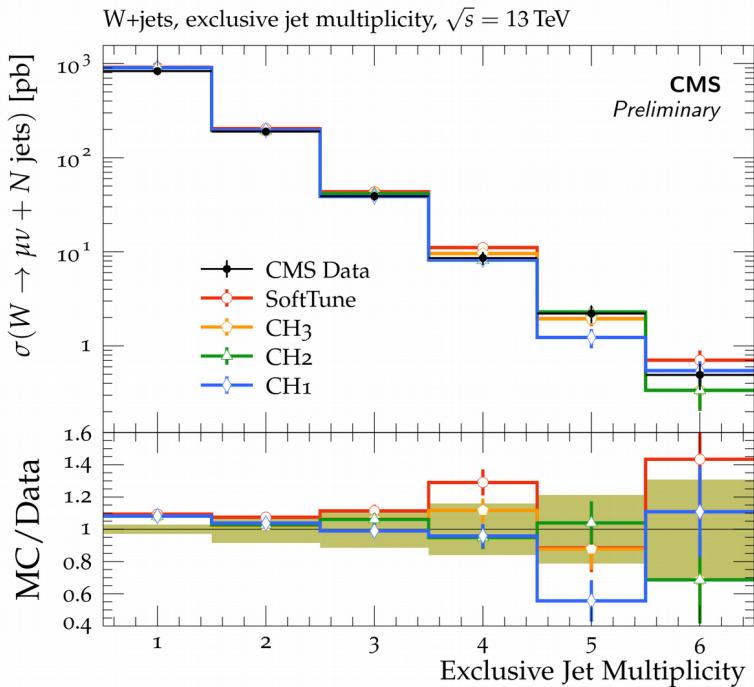


- **NLO ME** calculation with NNLO PDF set, Powheg + Herwig 7
- **Kinematic properties well described**

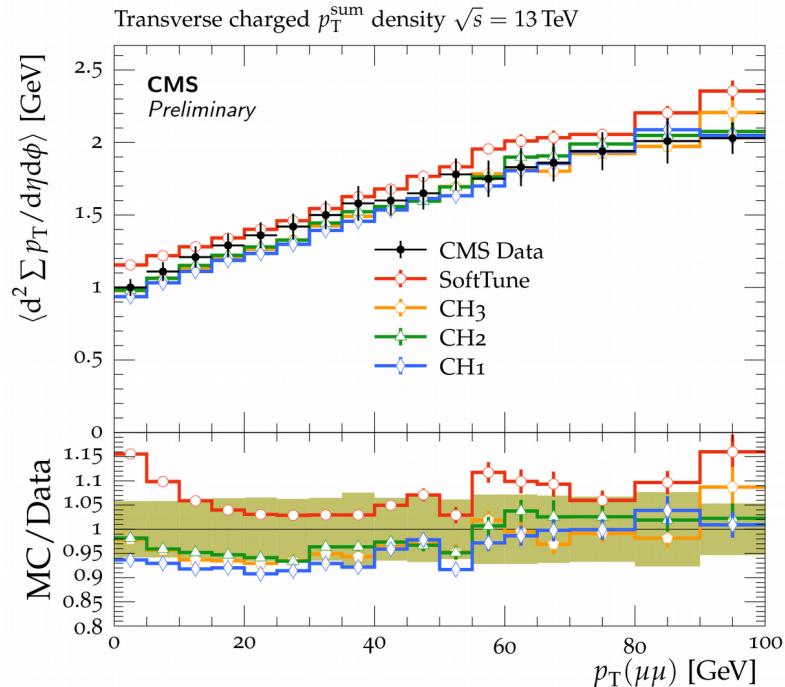
- **CH tunes underestimate high jet multiplicity:** lower α_s than SoftTune, but scale uncertainties expected to be large at high multiplicity

Example validations of different Herwig 7 tunes

W/Z + jets production



UE in Z + jets production



- **NLO ME** calculations up to 2 additional partons
MG5_aMC@NLO + Herwig 7 [FxFx]
- ME (PS) dominates at low (high) jet multiplicity
- **Similar description by all tunes**

- Checks UE description at higher scales than MB data
- **CH tunes describe data well**

Summary

CMS extracted and validated sets of tunes for the UE simulation,
by fitting predictions to UE-sensitive observables in MB events at various \sqrt{s}

Pythia 8 [GEN-17-001](#)

- New tunes improve on older Pythia tunes extracted at lower \sqrt{s}
- Predictions based on higher-order PDF shown to give reliable description of MB and UE measurements
- CP5 tune provides good overall description, most used in CMS

Herwig 7 [GEN-19-001](#)



- First dedicated Herwig 7 tunes with CMS data
- Derived tunes with different assumptions on PDF and α_s in UE
- Good overall description; CH3 tune will be used in CMS

Summary

CMS extracted and validated sets of tunes for the UE simulation,
by fitting predictions to UE-sensitive observables in MB events at various \sqrt{s}

Pythia 8 [GEN-17-001](#)

- New tunes improve on older Pythia tunes extracted at lower \sqrt{s}
- Predictions based on higher-order PDF shown to give reliable description of MB and UE measurements
- CP5 tune provides good overall description, most used in CMS

Herwig 7 [GEN-19-001](#)



- First dedicated Herwig 7 tunes with CMS data
- Derived tunes with different assumptions on PDF and α_s in UE
- Good overall description; CH3 tune will be used in CMS

UE simulation from new tunes **interfaced with higher-order and multileg ME** provide good description

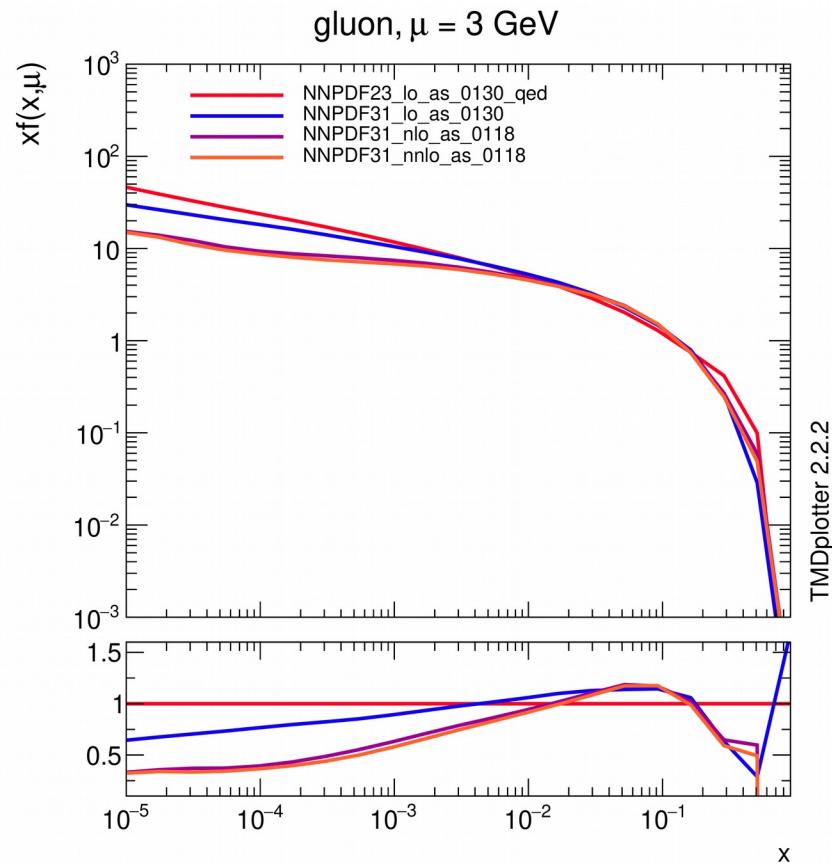
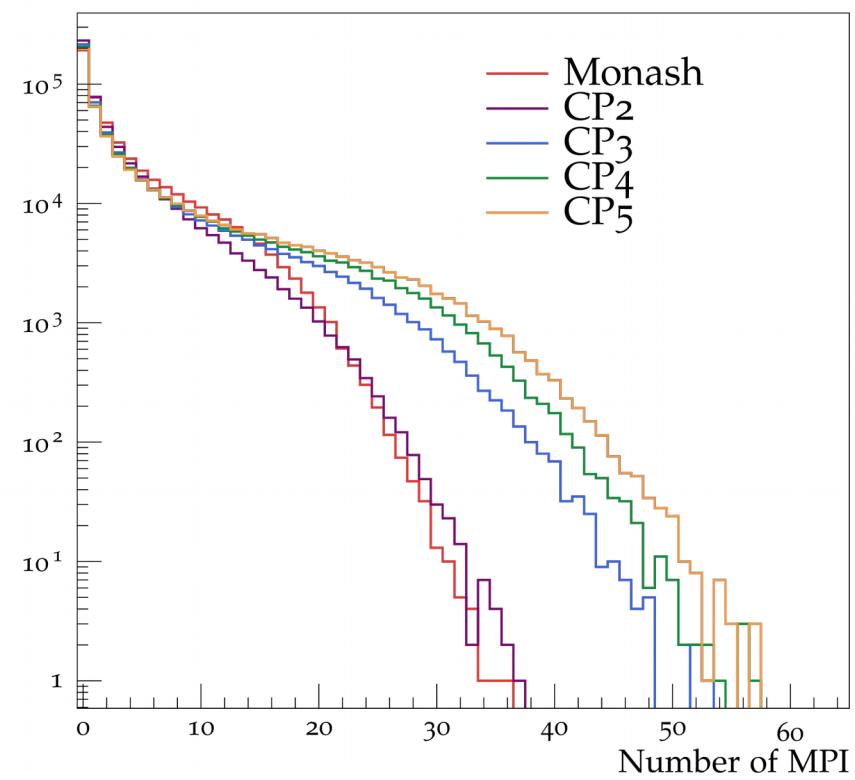
Extensive validation: event-shape observables, MB & UE simulation not used in tuning, multijet events, top quark pairs, W/Z production, double-parton scattering

Tunes + uncertainties widely used in Run 2 CMS measurements and searches

Backup

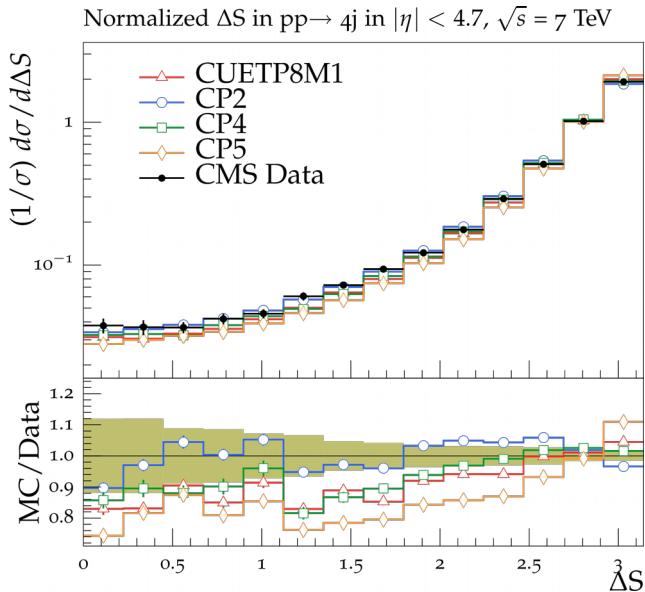
Number of MPI and gluon PDF in CP tunes

Number of MPI for various tunes

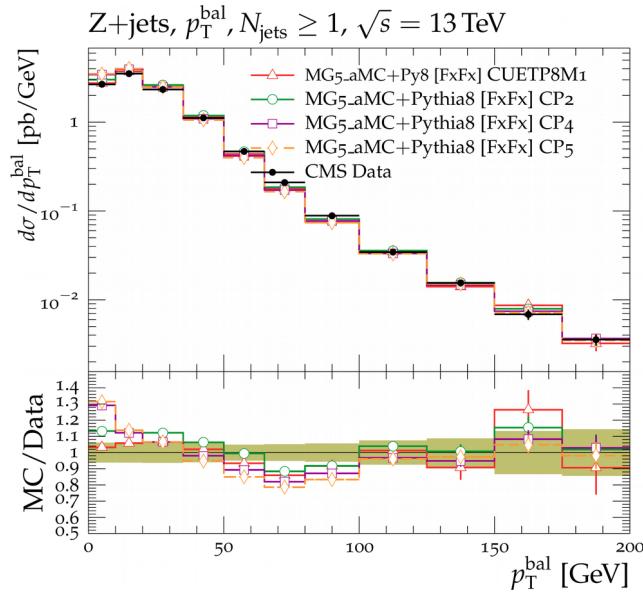


More validations of different Pythia 8 tunes

Double-parton scattering (DPS)



W/Z + jets production



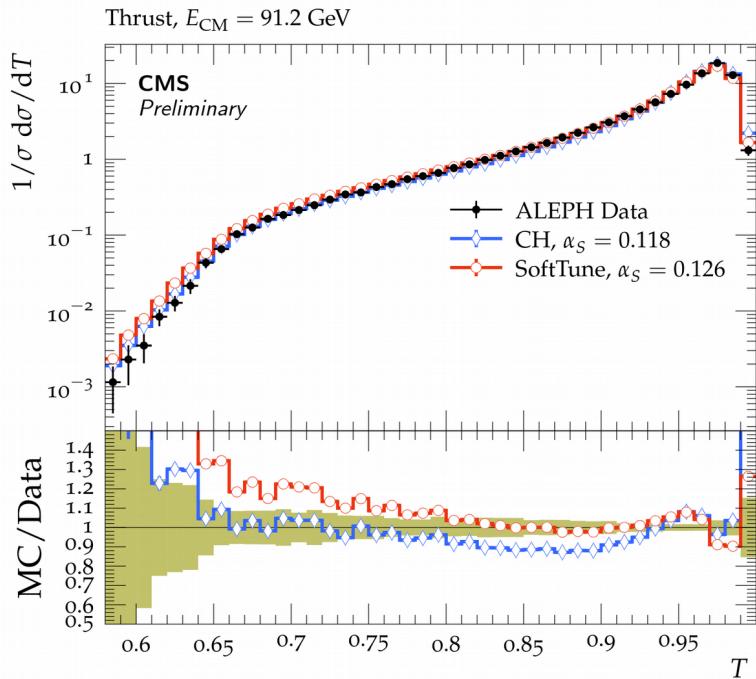
- **Topology in transverse plane** sensitive to contribution from DPS
- ΔS observable relates production planes of the hard and soft jet pairs

$$\Delta S = \arccos \left(\frac{\vec{p}_{T,1} \cdot \vec{p}_{T,2}}{|\vec{p}_{T,1}| |\vec{p}_{T,2}|} \right)$$
- **Best described by CP2 tune** due to different amount of MPI

- MG5_aMC@NLO + Pythia 8 at **L₀** [MLM] or **NLO** [FxFx]
- **Some observables** such as p_T balance between Z boson and jet, and $p_T(Z)$, **best described by CP2**

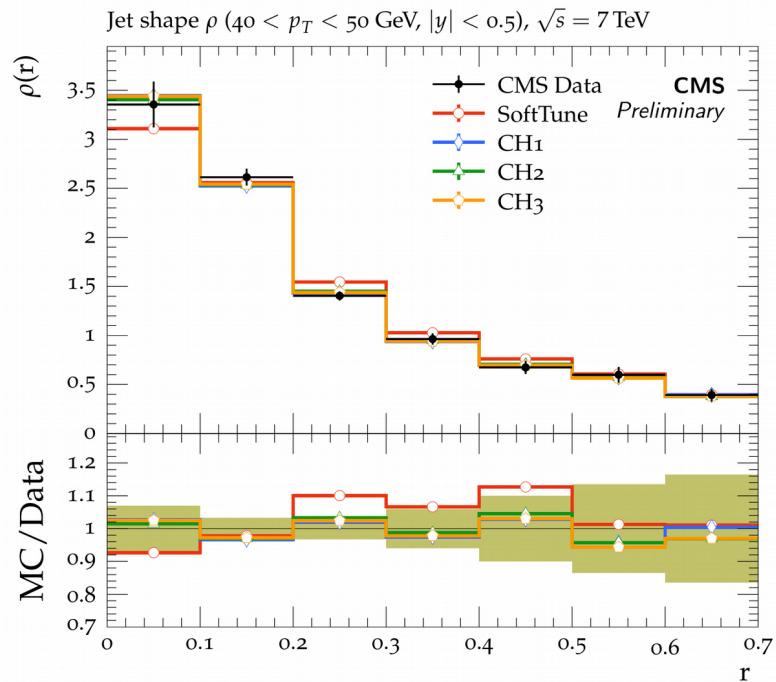
More validations of different Herwig 7 tunes

LEP event shape observables



- CH tunes underestimate number of events with $0.8 < T < 0.95$, but SoftTune predicts too many isotropic events (low values $T < 0.8$)

Inclusive jet production



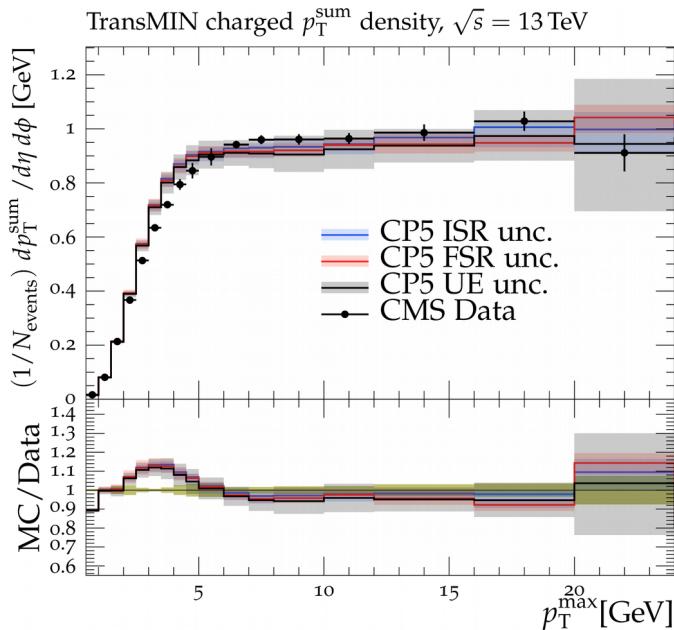
- Probing jet substructure
- Jet shape $\rho(r)$:** average fraction of p_T of jet constituents contained inside annulus with inner (outer) radius $r-0.1$ ($r+0.1$)
- CH tunes describe data well

UE tune uncertainties

Two tune variations (“up” and “down”)

- Constructed from $2n$ eigentunes (n = number of tuning parameters)
 - Summing differences in quadrature in each bin
 - Variations fitted to obtain the two sets of tune parameters
- Simplifies uncertainty estimation for CMS analyses while still replicating combination of all $2n$ eigentunes

Pythia 8



Herwig 7

