Event generation at CMS

May 2nd, 2019

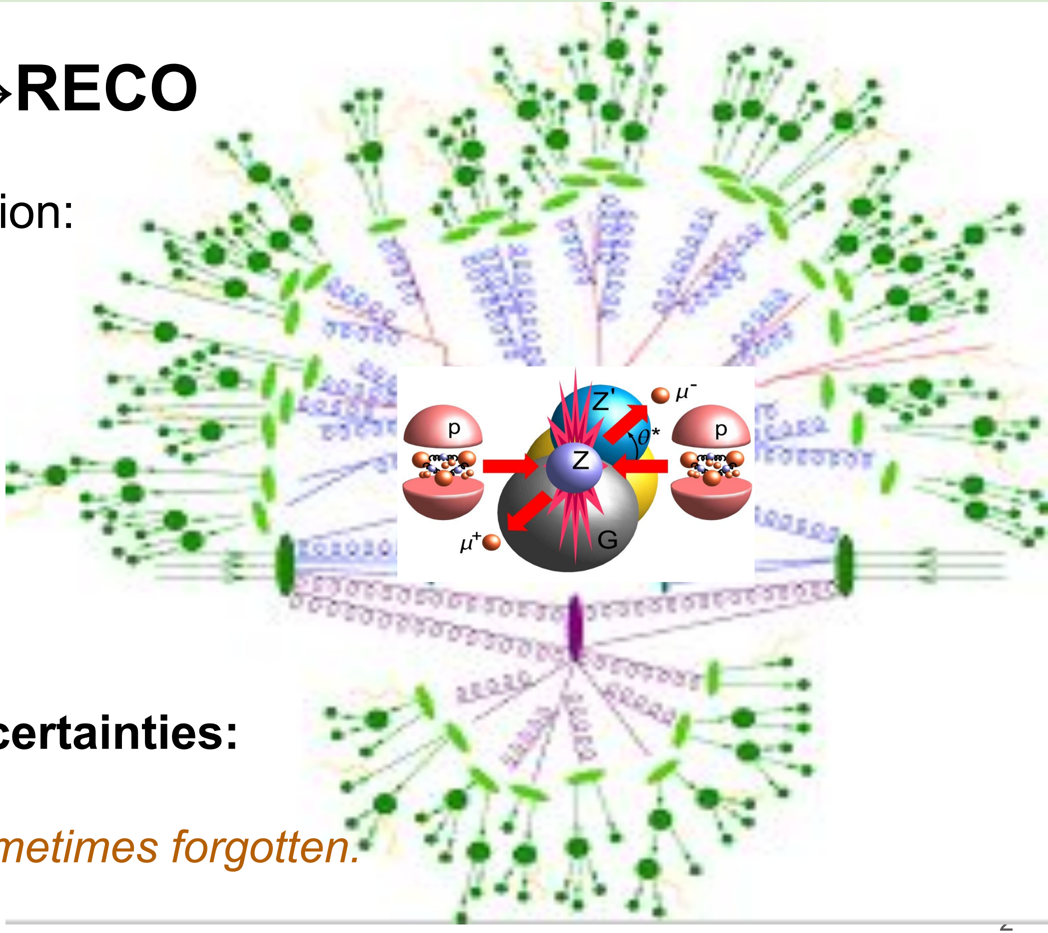
Qiang Li (Peking University)

Game of Flavours - CMS Heavy flavour tagging workshop 2019

CMS MC Simulation

(LHE→) GEN→ SIM→ DIGI→RECO

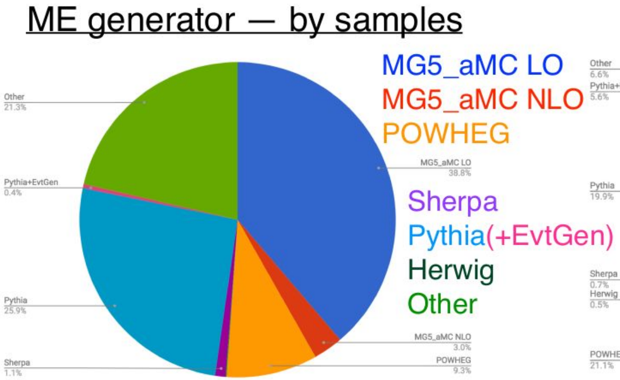
– Hard process/Matrix Element generation:
    parton level, perturbative QCD
– Parton Shower/Hadronization:
    QCD/QED emissions to a low scale,
    Produces hadrons from QCD partons
– Multiple Parton Interaction
– Detector Simulation and Digitization
– Reconstruction

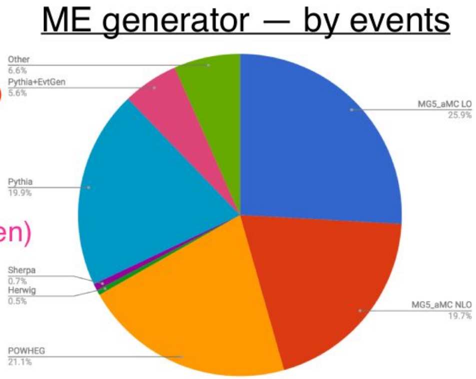


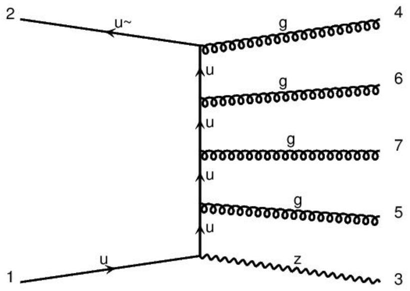
Factorised approach may lead to uncertainties:
PDF, ME, "Tune" and PS, …
→ *In most cases they are relevant but sometimes forgotten.*

Generator usage in CMS

based on 2016 MC campaign







- Multi-leg LO and NLO consistently matched to the parton shower
  - LO: Z+0/1/2/3/4 Jets

    Most commonly used in CMS: MG5_aMC@NLO+Pythia8 with MLM matching

    Most complex process up to 4 additional jets
  - NLO TTbar+0/1/2 Jets

    Most commonly used in CMS: MG5_aMC+Pythia8 with FxFx merging

    Most complex process up to 2 additional jets at NLO.
- For signal, NNLO+PS

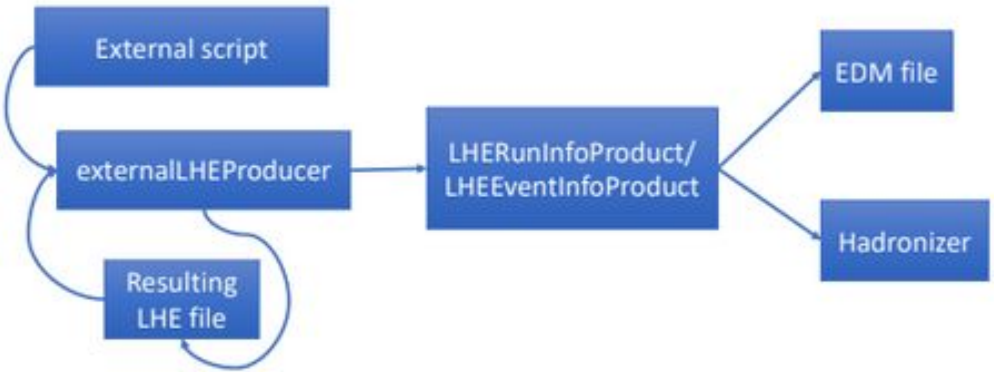
  POWHEG: MINLO_NNLOPS

  CMS HWW reweight the nominal signal to this one

Gridpack

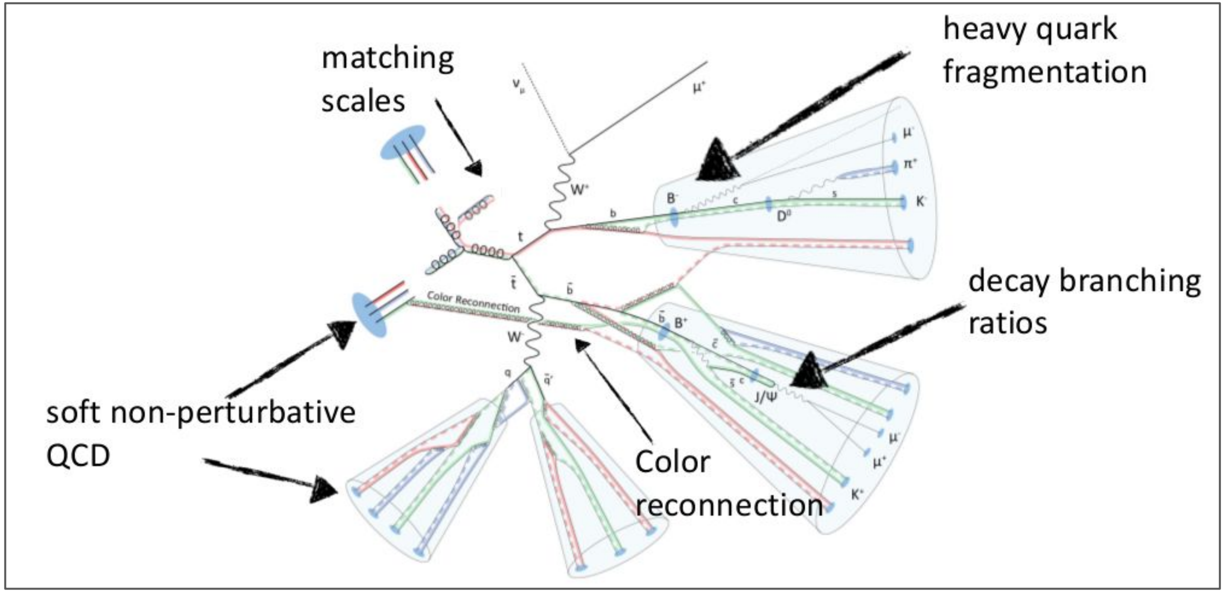
Gridpack (LHE) → GEN→ SIM→ DIGI→ RECO

- Pre-generated and compiled code with initial phase space integration results stored in a tarball (with fixed model/run parameters) → Gridpack
- Gridpack placed in CVMFS and accessed by remote jobs
- Scripts maintained for all major generators to produce gridpack tarballs LSF / Condor / CMS-Connect (grid-like condor jobs using CMS Global Pool)



Showering and Hadronization

- Pythia 8.226~ default; Herwig++ replaced by Herwig7.
- Fragment settings depend on Matrix Element
  - LO, NLO; MLM, FxFx matching/merging; POWHEG emission vetoing
- Pythia 8.240 integrated in CMS recently:
  - NLO shower DIRE, VINCIA;   Dipolerecoil for better description on VBS



More weights included than last year

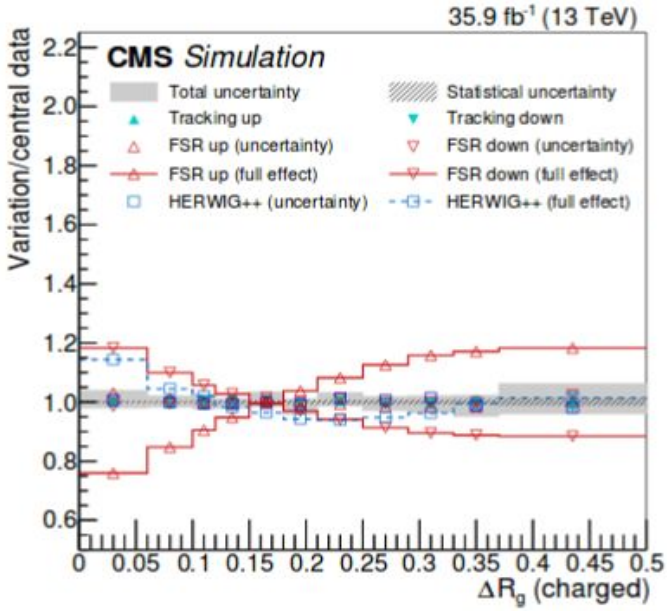
Showering and Hadronization

- Parton Shower Weights, Tune Variations etc included

$$(\text{ISR, FSR}) \otimes (\mu_R, \text{cNS}) \otimes (g \to gg,\ g \to q\bar{q},\ q \to qg,\ b/t \to b/t + g) \otimes (\text{up, down})$$

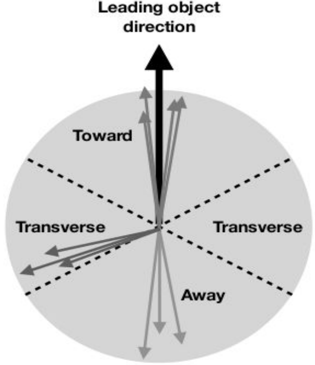
Source	Handle	Weights	Variation	Note/Reference	Dedicated studies
Shower scales	ISR scale (SpaceShower:renormMultFac) FSR scale (TimeShower:renormMultFac)	YES	0.5-2.0 0.5-2.0	FSR variations can be scaled down by √2 from LEP	TOP-15-011, TOP-16-021 TOP-17-13, TOP-17-015, ...
ME-PS Matching	hdamp	No	hdamp=1.58m_t +0.66-0.59 m_t	see TOP-16-021	Starting scale variations for MG5_aMC@NLO
Soft QCD	UE parameters	YES	UE tune up/down	See TOP-16-021 MPI & CR strength doesn't affect resonance decays	TOP-17-015 GEN-17-001
Color reconnection (odd clusters)	MPI based, QCD-inspired, gluon move	No	different models	CR affecting resonance decays	TOP-17-13, TOP-17-015
Fragmentation	momentum transfer from the b-quark to the B hadron: x_b=p_T(B)/p_T(b-jet)	YES	Vary Bower-Lund parameter within uncertainties from LEP/SLD fits	see TOP-16-022 (re-weight x_b)	
Flavor response/ hadronization	Pythia vs Herwig	No	Vary the JES independently per flavour for light, g, c, b.		
Decay tables	B semi-leptonic BR	YES	vary semileptonic BR +0.77%/-0.45%	re-weight the fraction of semi-leptonic b jets by the PDG values (scale Λ_b to match PDG)	

CMS: Jet substructure in ttbar events at 13 TeV (2016)

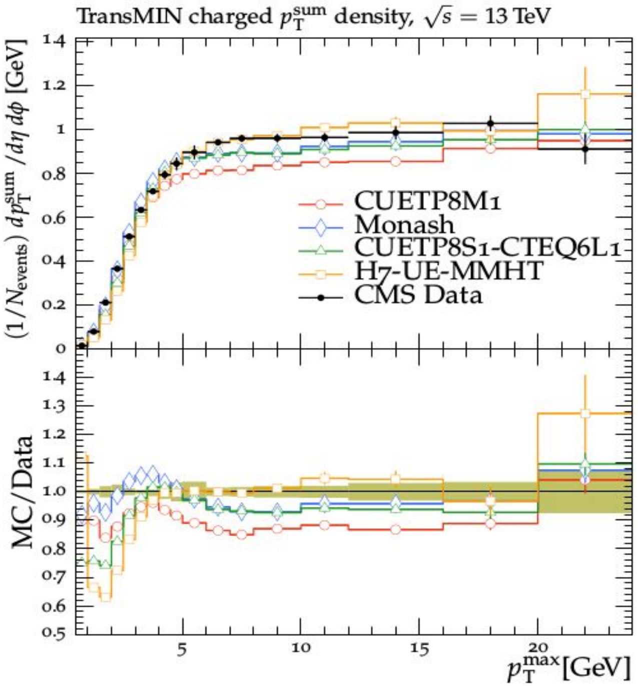


angle between the groomed subjets

MC Tuning: CUETP8M1 Tune



Until 2017 analyses (except 2016 ttbar), **Pythia8 CUETP8M1 tune** [EPJC 76 (2016) 155] based on the **Monash** tune was used. Fitting MPI energy dependence parameters to UE data @sqrt(s) = 0.9, 1.96 & 7 TeV



TOP-16-021

- α_s and shower parameters kept as in Monash → $\alpha_s^{ISR/FSR}$=0.1365 despite the prefered values of 0.130 in LO and 0.118 in NLO matrix elements/ PDF sets.
  - α_s^{FSR} in Monash → by fitting Pythia8 predictions to LEP event shape measurements and α_s^{ISR} is just assumed to be the same as α_s^{FSR}.
  - α_s^{MPI} = 0.130 set to the value prefered in the LO PDF set.
- Revisited the shower parameters
  - Starting from parton shower in ttbar events → CUETP8M2T4 tune.
  - Using a NNLO PDF set in PS → CP5 (and CP0-4 tunes).

CUETP8M1 does not describe well the central values of 13TeV data

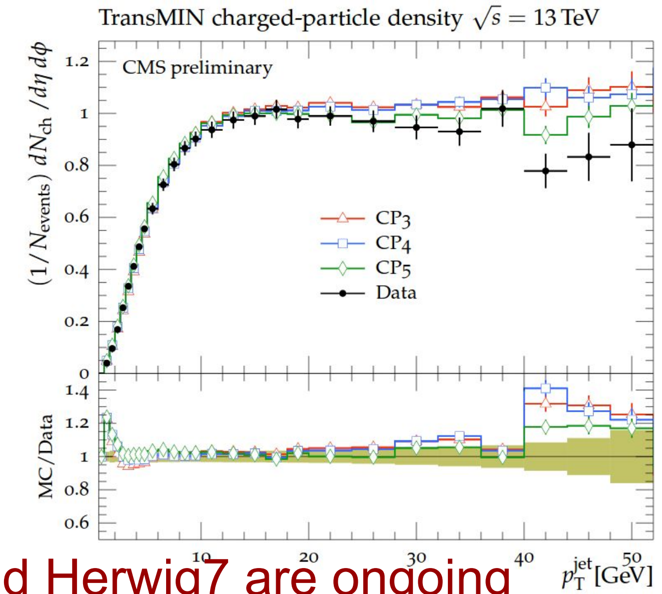
MC Tuning: CPX Tune

arXiv:1903.12179

- **First CMS tune with 13 TeV LHC data**
- **Match PDF and αs in the PS and in the ME.**
  - PYTHIA tunes are mostly based on LO PDFs.
  - Sherpa tunes are based on NNLO PDFs.
  - HERWIG7 provide tunes based on NLO PDFs
- **CP5 with NNPDF3.1NNLO as default for 2017/2018 MC productions**

- CP1: NNPDF3.1 LO (α_S =0.130)
- CP2: NNPDF3.1 LO (α_S =0.130)
- CP3: NNPDF3.1 NLO (α_S =0.118)
- CP4: NNPDF3.1 NLO (α_S =0.118)
- CP5: NNPDF3.1 NNLO (α_S =0.118)

PYTHIA8 parameter	CP1	CP2	
PDF Set	NNPDF3.1 LO	NNPDF3.1 LO	Fixed inputs
$\alpha_S(m_Z)$	0.130	0.130	
SpaceShower:rapidityOrder	off	off	
MultipartonInteractions:EcmRef [GeV]	7000	7000	
$\alpha_S^{ISR}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_S^{FSR}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_S^{MPI}(m_Z)$ value/order	0.130/LO	0.130/LO	
$\alpha_S^{ME}(m_Z)$ value/order	0.130/LO	0.130/LO	
MultipartonInteractions:pT0Ref [GeV]	2.4	2.3	Fitted parameters
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	

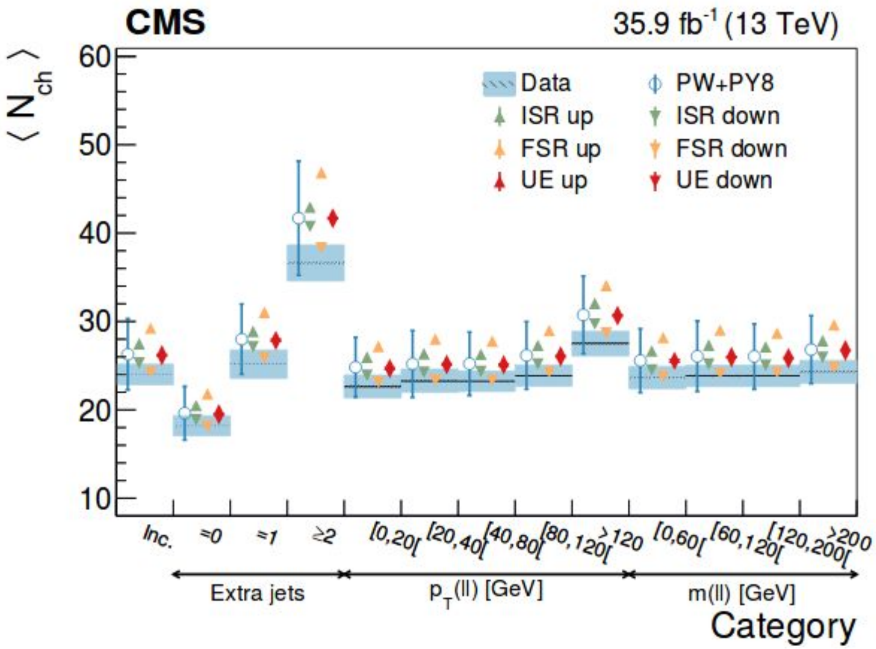


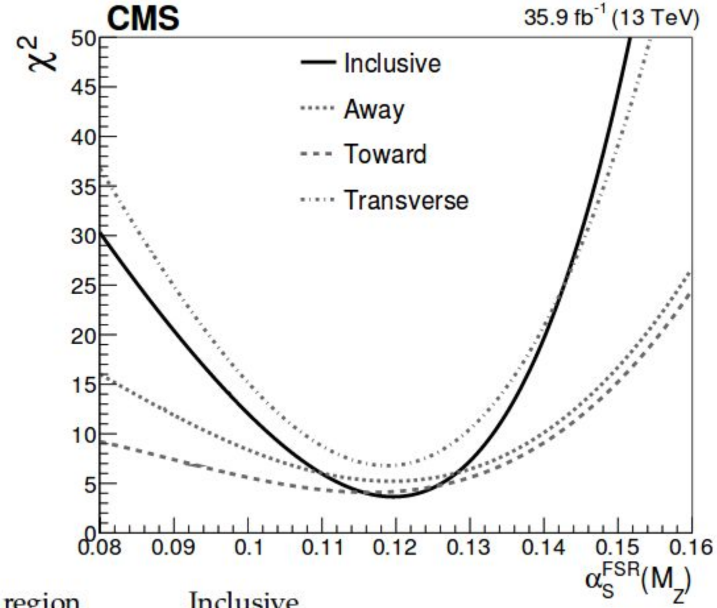
CMS Tunes with Pythia8 Color Reconnection and Herwig7 are ongoing

UE in TTbar

EPJC79(2019)123

- UE Measurement for the first time at a scale > 2mt
- Measurement unfolded to particle level
- Good agreement of POWHEG+PYTHIA8 with CUETP8M2T4 in UE





$p_T(\ell\ell)$ region	Inclusive
Best fit $\alpha_S^{FSR}(M_Z)$	0.120
68% CI	[-0.006,+0.006]
95% CI	[-0.013,+0.011]

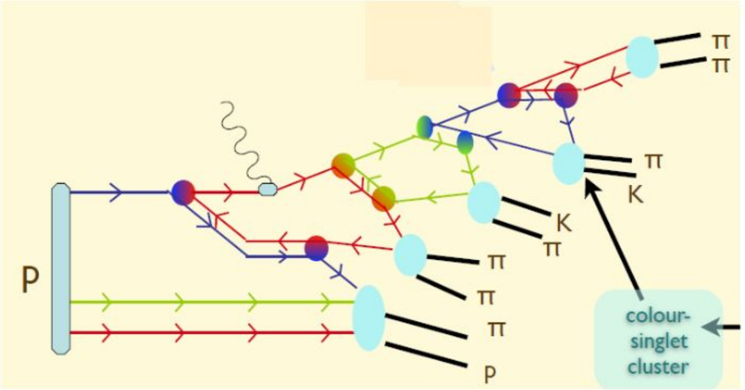
Top Mass: b jet/ps/Color-reconnection EPJC79 (2019) 313

- **Bowler–Lund fragmentation function** varied within uncertainties determined by the ALEPH and DELPHI
- Alternatively **Peterson fragmentation function**.
- **Semileptonic b hadron branching fraction**, varied within B0/B+decays measurement and uncertainties

$$m_t^{hyb} = 172.26 \pm 0.07\,(\text{stat+JSF}) \pm 0.61\,(\text{syst})\,\text{GeV}$$

Top quark mass from ttbar fully hadronic (2016)

	2D δm_t^{2D} [GeV]	2D δJSF^{2D} [%]	1D δm_t^{1D} [GeV]	hybrid δm_t^{hyb} [GeV]	hybrid δJSF^{hyb} [%]
b jet modeling (quad. sum)	0.09	0.0	0.09	0.09	0.0
– b frag. Bowler–Lund	−0.07	0.0	−0.07	−0.07	0.0
– b frag. Peterson	−0.05	0.0	−0.04	−0.05	0.0
– semileptonic b hadron decays	−0.03	0.0	−0.03	−0.03	0.0
PDF	0.01	0.0	0.01	0.01	0.0
Ren. and fact. scales	0.05	0.0	0.04	0.04	0.0
ME/PS matching	$+0.32 \pm 0.20$	−0.3	-0.05 ± 0.14	$+0.24 \pm 0.18$	−0.2
ISR PS scale	$+0.17 \pm 0.17$	−0.2	$+0.13 \pm 0.12$	$+0.12 \pm 0.14$	−0.1
FSR PS scale	$+0.22 \pm 0.12$	−0.2	$+0.11 \pm 0.08$	$+0.18 \pm 0.11$	−0.1
Top quark p_T	+0.03	0.0	+0.02	+0.03	0.0
Underlying event	$+0.16 \pm 0.19$	−0.3	-0.07 ± 0.14	$+0.10 \pm 0.17$	−0.2
Early resonance decays	$+0.02 \pm 0.28$	+0.4	$+0.38 \pm 0.19$	$+0.13 \pm 0.24$	+0.3
CR modeling (max. shift)	$+0.41 \pm 0.29$	−0.4	-0.43 ± 0.20	-0.36 ± 0.25	−0.3
– "gluon move" (ERD on)	$+0.41 \pm 0.29$	−0.4	$+0.10 \pm 0.20$	$+0.32 \pm 0.25$	−0.3
– "QCD inspired" (ERD on)	-0.32 ± 0.29	−0.1	-0.43 ± 0.20	-0.36 ± 0.25	−0.1
Total systematic	0.81	0.9	1.03	0.70	0.7
Statistical (expected)	0.21	0.2	0.16	0.20	0.1
Total (expected)	0.83	0.9	1.04	0.72	0.7



Michelangelo L. Mangano
CLASHEP 2017

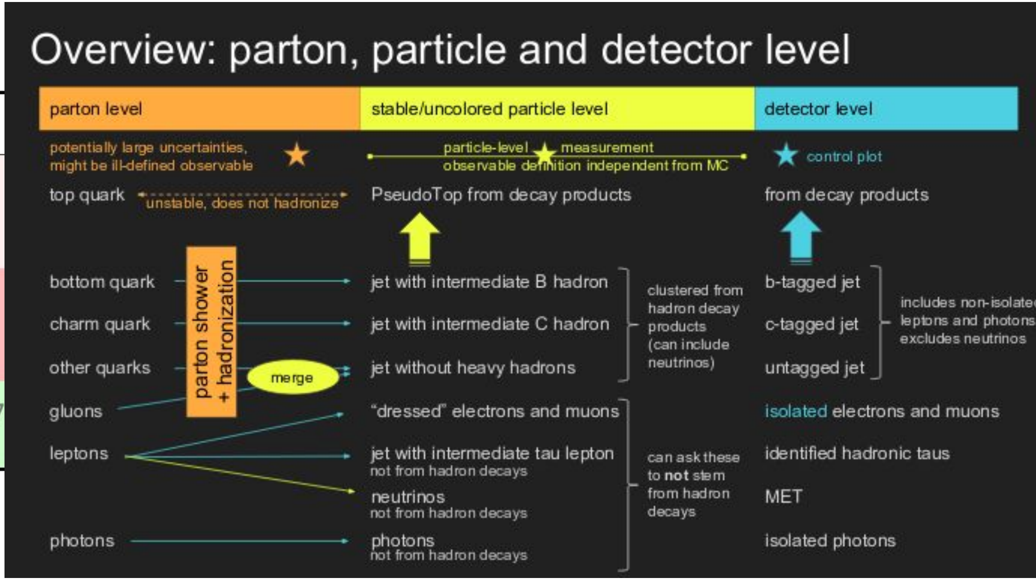
CR: MPI or QCD based, Gluon-Move

Rivet & Particle-Level Objects [CERN-CMS-NOTE-2017-004]

- **Rivet preserves complete particle-level analyses in form of code.**
  Provides a lot of easy-to-use analysis “projections”, e.g.
  FastJets with ghost bottom/charm/τ tagging automatically included,
  no distinction between light quarks and gluons (but g→bb/cc possible)
- **Recently implemented Rivet definitions in CMSSW module:**
  ParticleLevelProducer, writing a Rivet plugin of the analysis becomes fast and simple.

Key	ALICE	ATLAS	CMS	LHCb	B-factories	HERA	Other
Rivet wanted (total):	200	254	345	161	1498	450	2276
Rivet REALLY wanted:	35	40	74	10	2	14	2
Rivet provided:	20/220 = 9%	145/399 = 36%	74/419 = 18%	11/172 = 6%	14/1512 = 1%	4/454 = 1%	41/2317 = 2%

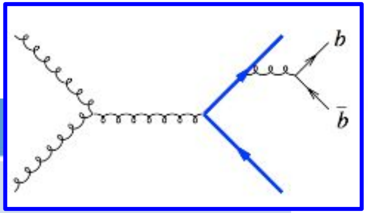
Markus Seidel



ttbb/ttjj

[PLB 776 (2018) 355]

Crucial bkg to ttH (H⟶bb)
Dilepton events with≥4 jets,≥2 b–tagged jets
Signal extracted by fitting to b-discriminators
Categorized with Particle level objects
Main uncertainties:
    b/mistag efficiency, modelling



Event generator	Parton shower
POWHEG (v2)	Pythia 8
POWHEG	Herwig++
MadGraph5_aMC@NLO	Pythia 8 with FxFx ($t\bar{t}$ + 0, 1, 2j)
MadGraph5 (LO)	Pythia 8 with MLM ($t\bar{t}$ + 0, 1, 2, 3j)
MadGraph5_aMC@NLO	Herwig++

$t\bar{t}$+heavy flavour events are extracted from inclusive $t\bar{t}$ sample

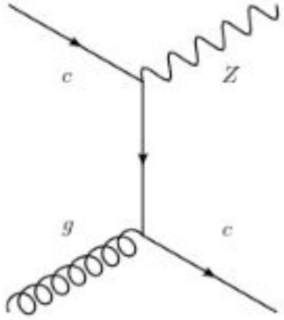
Source	$\sigma_{t\bar{t}b\bar{b}}$	$\sigma_{t\bar{t}jj}$	$\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$
b tag (b quark flavour)	19	4.7	19
b tag (c quark flavour)	14	1.3	14
b tag (light flavour)	14	9.8	9.7
JES & JER	7.8	7.4	2.6
Ratio of $t\bar{t}b\bar{b}$ and $t\bar{t}bj$	2.6	0.5	2.6
Background modelling	3.8	3.5	1.6
$t\bar{t}c\bar{c}$ fraction in the fit	5.2	1.9	4.8
Lepton trigger/identification	3.0	3.0	0
Pileup	0.4	<0.1	0.4
MC generator	9.4	6.2	3.0
μ_F and μ_R scale	2.0	2.0	1.0
scale in PS	13	9.9	10
PDFs	0.5	0.5	<0.1
Efficiency ($t\bar{t}c\bar{c}$ fraction)	0	1.3	1.3
Jet multiplicity modelling	5.0	5.0	5.0
Simulation (statistical)	1.5	1.5	1.5
Top quark p_T modelling	0.8	0.3	0.5
Integrated Luminosity	2.3	2.3	0
Total uncertainty	34	19	28

Phase space		$\sigma_{t\bar{t}b\bar{b}}$ [pb]	$\sigma_{t\bar{t}jj}$ [pb]	$\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$
Visible	Measurement	$0.088 \pm 0.012 \pm 0.029$	$3.7 \pm 0.1 \pm 0.7$	$0.024 \pm 0.003 \pm 0.007$
	SM (POWHEG)	0.070 ± 0.009	5.1 ± 0.5	0.014 ± 0.001

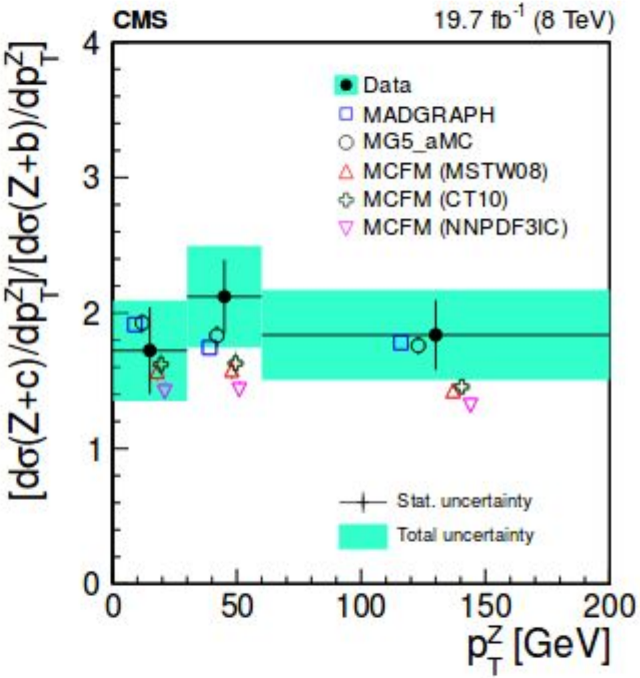
Z+c: PDF, decay

EPJC 78 (2018) 287

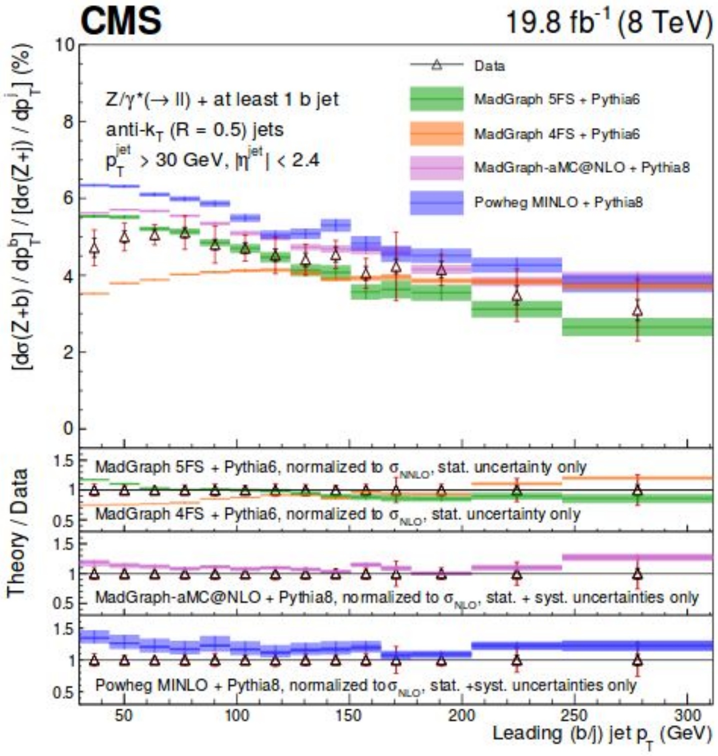
- Jets with HF quark content are identified through
  - (1) the semileptonic decay of c/b flavoured hadrons with a muon (muon-inside-a-jet) in the final state
  - (2) using exclusive hadronic decays of charm hadrons. ($D^{\pm}$, $D^{*}(2010)^{\pm}$)
- Events classified as Z+b, Z+c, or Z+light flavour
  - according to the flavour of the generator-level jets
- **Decay Br of e.g. Br($D^{\pm} \rightarrow K^{\mp}\pi^{\pm}\pi^{\pm}$) in PYTHIA:**
  - scaled in order to match the experimental values
- Fit to SV mass or JP discriminant

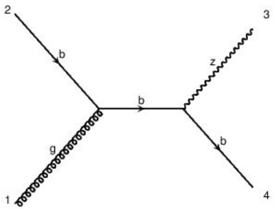


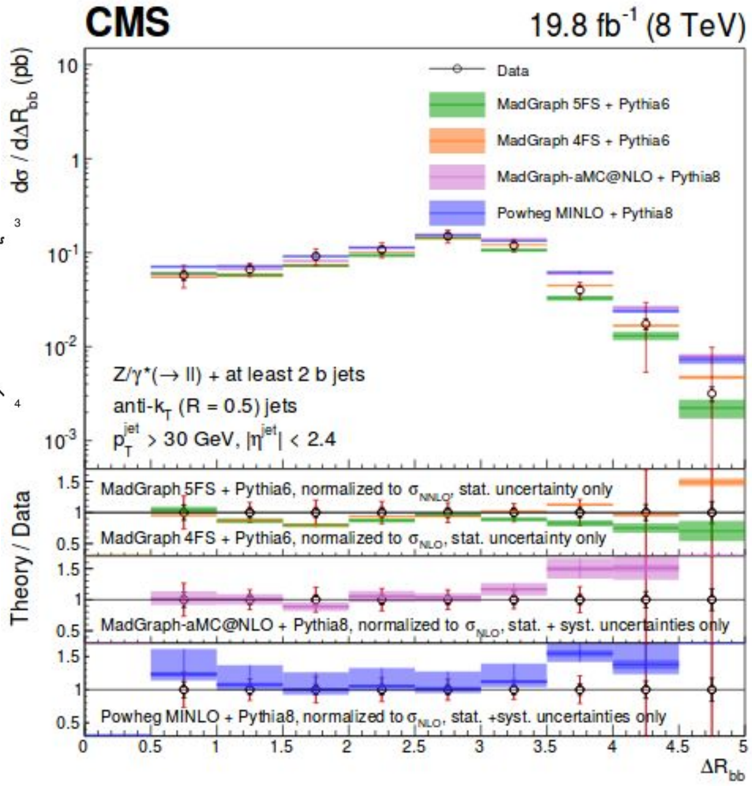
1. Fixed-order matrix elements matched to parton shower describe the data,both at LO and NLO;
2. MCFM fixed order NLO underestimate the cross sections;
3. **No big difference when using PDFs with and without intrinsic charm.**



Z+b/bb: 4f/5f, g→bb [EPJC 77 (2017) 751]







- Sizable differences between 4FS and 5FS at low pT.
- 5FS closer to data.

Gluon splitting dominated at small ΔR_{bb} region; Reasonable agreement

Searches: LO vs NLO

Relevant phase space are in tail of N(jets), VpT..., challenging to model.

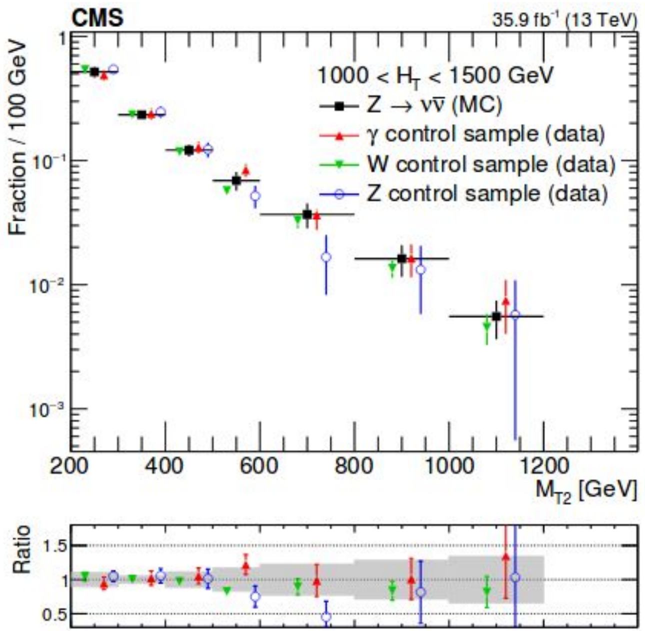
NLO vs LO: NLO more accurate, negative weights, statistically limited, CPU demanding; May rely on LO samples and correction factors from data

- Negative weights strongly reduce statistical power
- For weighted events w_i, effective events N_{eff} for fraction of negative weights f:
  $$N_{eff} = \frac{(\sum_i w_i)^2}{\sum w_i^2} = N(1-2f)^2$$
- for 35% negative weights (common at for high jet-mulitplicity/ high pt)
  $\Longrightarrow$ 9% effective events compared to $w_i = 1$

V(DATA-others, CR)*V(MC, SR)/V(MC, CR)

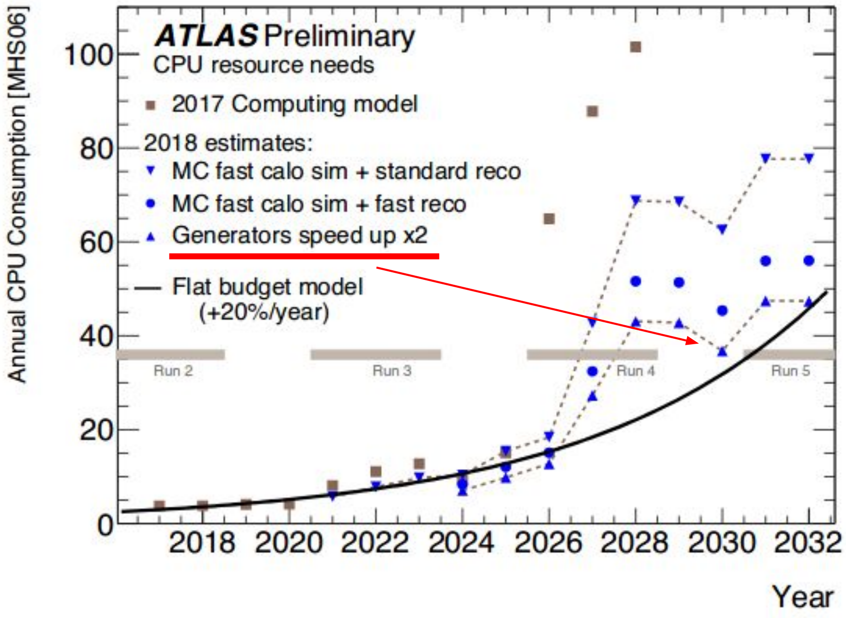
Examples of CR —> SR translation

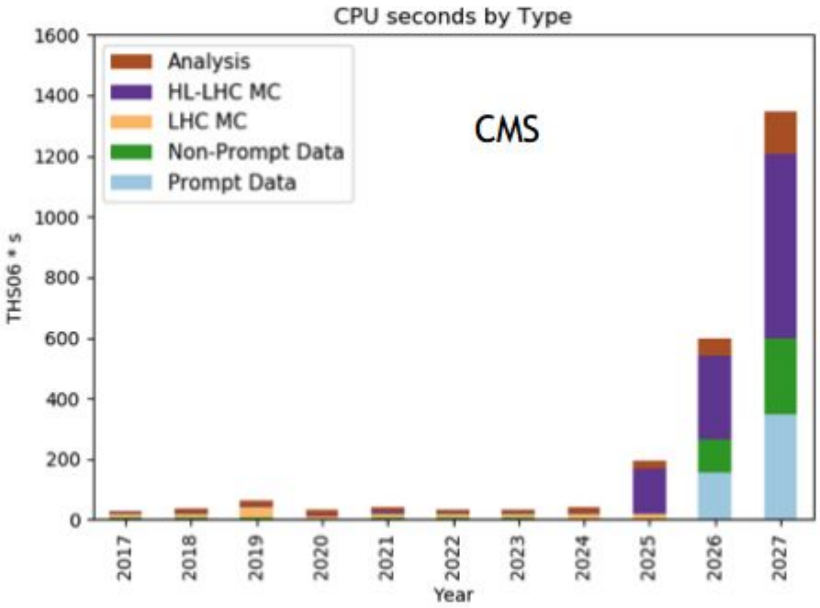
Final state	Background	Control sample	Features of extrapolation
0(1)-lepton + jets + MET	W+jets/ttbar with a "lost" lepton	1(2)-lepton	Correct lepton ID inefficiencies in MC to data, generally rely on MC to describe lepton acceptance
0-lepton + jets + MET	Z(vv)+jets	Z(ll)+jets or γ+jets	Hadronic recoil (MET excluding leptons or photon) in CR proxy for MET in SR, differences between Z and γ



SUSY MT2
https://arxiv.org/abs/1705.04650

Generator upgrade





- We are OFF by ~5x on CPU power when considering Moore's law
- HL-LHC salvation will come from software improvements, not from hardware

Elizabeth Sexton-Kennedy

HSF and Generator Workshop

The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing internationally.
Community White Paper: summarising R&D in a variety of technical areas for HEP Software and Computing

HEP Software Foundation

Physics Event Generator Computing Workshop
26 Nov 2018, 09:00 → 28 Nov 2018, 18:00 Europe/Zurich
4-3-006 - TH Conference Room (CERN)

Goals of this workshop:

○ Identify the most crucial areas for technical improvements to the generators used by the experiments.
○ Define a programme of work that can be used to attract investment in these technical areas, aiming to have software engineers who can work together with the generator authors.
○ Identify ways of making new theoretical advances easier to implement in a computationally efficient way.

Thanks

CMSDAS2019

One Week before CMSWeek in Thailand
CMS Students are welcome to join this school (registration will open soon)
Exercises are being organized. Facilitators/Contributions are welcome!

CMS Data Analysis School 2019 - Beijing

9-13 December 2019
Physics Department - Peking University
Asia/Hong_Kong timezone

Search...

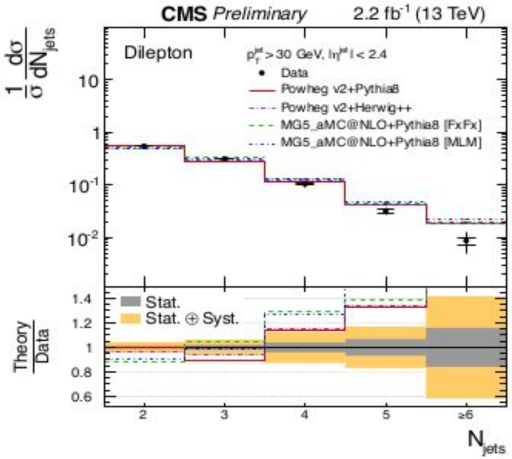
- Overview
- Scientific Program
- Timetable
- Local Organisation
- List of registrants
- CMSDAS Twiki
- List of Exercises and Facilitators
- Pre-exercises/instructions
- Assignment: Long Exercises

The CMS Data analysis school (CMSDAS) is the official school that CMS organize every year in US, in Europe and in Asia to train students, Ph.D and young post-docs for the physics analysis. It consists of two days of short exercises about physics objects reconstruction and identification and 2.5 days of long exercises about physics analysis in preparation for the Run2 data taking of the LHC. The school is a fundamental step for training new people and create the good network for people doing data analysis. Both the short and long exercises are designed and facilitated by teams of 2-3 CMS experts. Some long exercises go beyond the current state of the art of the corresponding CMS analysis. Thus providing opportunities to get plugged in to a CMS measurement through a long exercise at the school and become a part of the CMS measurement team on the paper. A prize will be assigned for the "Best Analysis Team" at the end of the school.

This year the Asian edition will be hosted in Beijing at the Physics Department of the Peking University.

Organized by **Qiang Li** (Chair)

MC Tuning: CUETP8M2T4 Tune [TOP-16-021]



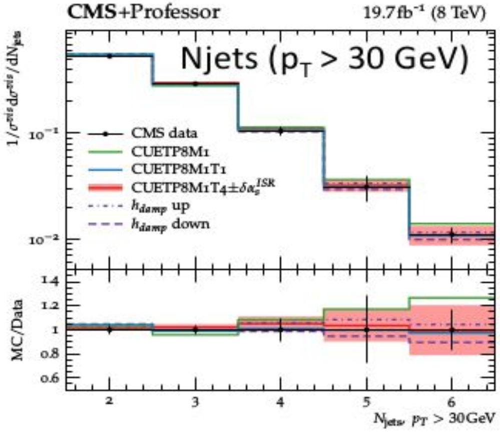


Table 3: The parameters of the old tune (CUETP8M1) and the new tune (CUETP8M2T4). The parameter (in italic) relative to the energy dependence of the partonic cross section cutoff (ECMPOW) is fixed to the value of CUETP8M1 in the fit of the new tune. In the new tune, the ISR α_s is fixed to the value extracted from the $t\bar{t}$ events described previously.

	CUETP8M1	CUETP8M2T4
Tune	pp 14	pp 14
Tune	ee 7	ee 7
MultipartonInteractions ecmPow	0.2521	0.2521
SpaceShower:alphaSvalue	0.1365	0.1108
PDF pSet LHAPDF6	NNPDF23_lo_qed_as_0130	NNPDF30_lo_as_0130
MultipartonInteractions:pT0Ref	2.40	2.20
MultipartonInteractions:expPow	1.6	1.6
ColourReconnection:range	1.8	6.6

CMS-PAS-TOP-16-021

Tune α_s^{ISR} using 8 TeV ttbar Njets and jet pT data —>

$$\alpha_s^{ISR} = 0.1108^{+0.0145}_{-0.0142}$$
$$h_{damp} = 1.581^{+0.658}_{-0.585} \times m_t$$

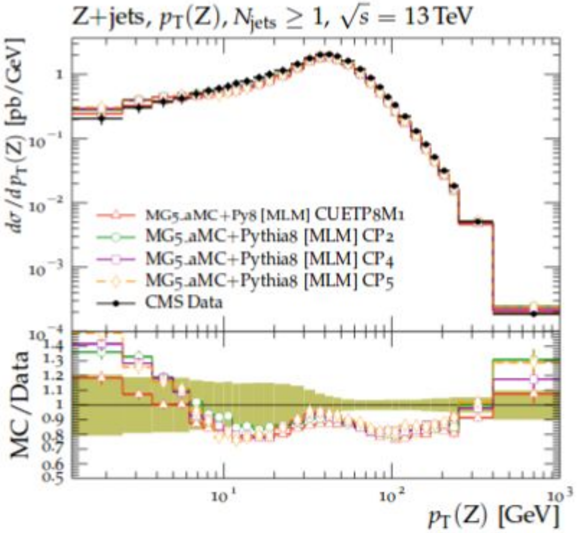
with **SpaceShowerRapidityOrdering=on** *(special care of options needed for the emissions produced by the PS)*

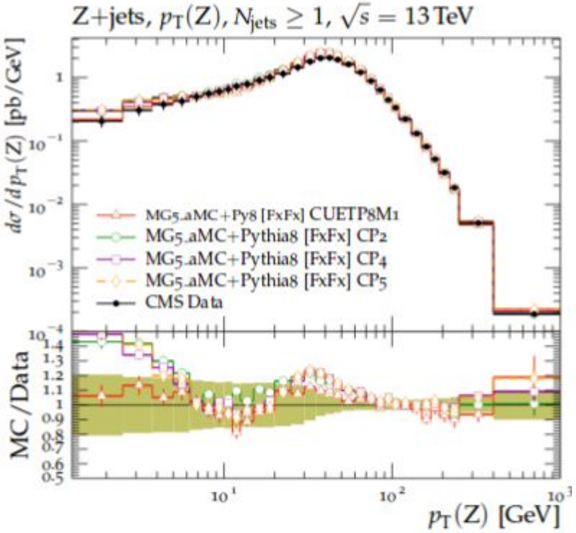
==> Significantly lower shower α_s cures the overshoot of CUETP8M1 at high jet multiplicities.
==> UE and min-bias are described better
==> POWHEG+PYTHIA8: generally consistent with data, with residual differences covered by theory uncertainties. arXiv:1803.0399

Efe Yazgan

MC Tuning: CPX Tune

arXiv:1903.12179





New Tunes based on higher-order PDF sets, interfacing with higher-order and multileg matrix element generators, such as POWHEG and MG5_aMC, gives a **reliable** description of observables measured in multijet final states, Drell-Yan, and top quark production processes

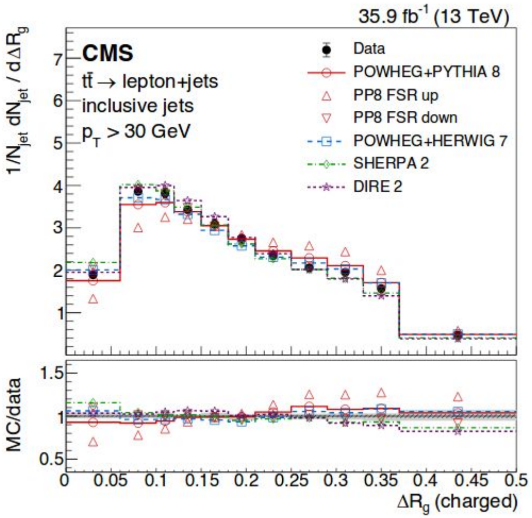
CMS Tunes with Pythia8 Color Reconnection and Herwig7 are ongoing

MC Tuning: jet substructure

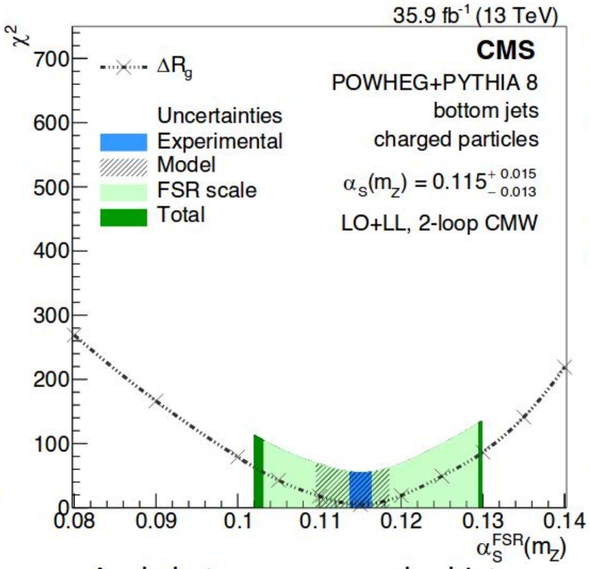
α_s^{FSR} from jet substructure in ttbar l+jets events

- Measured using charged+neutral and with only charged jet constituents (particle pT > 1 GeV).
- b, light, or gluon jet enriched samples.

PRD.98.092014



Angle between groomed subjets at particle level (correlated to jet width)



Angle between groomed subjets at particle level.

Pythia8:
CUETP8M2T4 for ttbar
CUETP8M1 for the rest.

None of the default tunes yield a good overall description of the data.

Powheg+Pythia8 with LO+LL, 2-loop :

$$\alpha_s^{FSR}(M_Z) = 0.115^{+0.015}_{-0.013}$$

Efe Yazgan DIS2019