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 $TT\text{bar}+0/1/2$ 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

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

<table>
<thead>
<tr>
<th>Source</th>
<th>Handle</th>
<th>Weights</th>
<th>Variation</th>
<th>Note/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shower scales</td>
<td>(SpaceShower:renormMultFac)</td>
<td>0.5-2.0</td>
<td></td>
<td>FSR variations can be scaled down by (\sqrt{2}) from LEP</td>
</tr>
<tr>
<td></td>
<td>(TimeShower:renormMultFac)</td>
<td>0.5-2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME-PS Matching</td>
<td>hdamp</td>
<td>No</td>
<td></td>
<td>see TOP-16-021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Starting scale variations for MGS, aM@NLO</td>
</tr>
<tr>
<td>Soft QCD</td>
<td>UE parameters</td>
<td>YES</td>
<td></td>
<td>See TOP-16-021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MPI &amp; CR strength doesn’t affect resonance decays</td>
</tr>
<tr>
<td>Color reconnection (odd clusters)</td>
<td>MPI based, QCD-inspired, gloun move</td>
<td>No</td>
<td>different models</td>
<td>CR affecting resonance decays</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragmentation</td>
<td>momentum transfer from the b-quark to the B hadron: (x_0 = \frac{p_B}{p_T(b\text{-jet})})</td>
<td>YES</td>
<td></td>
<td>Vary Bower-Lund parameter within uncertainties from LEP/SLD fits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>see TOP-16-022 (\text{re-weight } x_0)</td>
</tr>
<tr>
<td>Flavor response/hadronization</td>
<td>Pythia vs Herwig</td>
<td>No</td>
<td></td>
<td>Vary the JES independently per flavour for light, g, c, b.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decay tables</td>
<td>B semi-leptonic BR</td>
<td>YES</td>
<td></td>
<td>vary semi-leptonic BR +0.77%/-0.45%</td>
</tr>
</tbody>
</table>
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

- \(\alpha_s^{\text{ISR/FSR}}\)=0.1365 despite the preferred values of 0.130 in LO and 0.118 in NLO matrix elements/ PDF sets.
  - \(\alpha_s^{\text{FSR}}\) in Monash → by fitting Pythia8 predictions to LEP event shape measurements and \(\alpha_s^{\text{ISR}}\) just assumed to be the same as \(\alpha_s^{\text{FSR}}\).
  - \(\alpha_s^{\text{MPI}}\) = 0.130 set to the value preferred 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

- First CMS tune with 13 TeV LHC data
- Match PDF and $\alpha_s$ 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

<table>
<thead>
<tr>
<th>PYTHIA8 parameter</th>
<th>CP1</th>
<th>CP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF Set</td>
<td>NNPDF3.1 LO</td>
<td>NNPDF3.1 LO</td>
</tr>
<tr>
<td>$a_s(m_Z)$</td>
<td>0.130</td>
<td>0.130</td>
</tr>
<tr>
<td>SpaceShower:rapidityOrder</td>
<td>off</td>
<td>off</td>
</tr>
<tr>
<td>MultipartonInteractions:EcmRef [GeV]</td>
<td>7000</td>
<td>7000</td>
</tr>
<tr>
<td>$a_s^{\text{FR}}(m_Z)$ value/order</td>
<td>0.1365/LO</td>
<td>0.130/LO</td>
</tr>
<tr>
<td>$a_s^{\text{FR}}(m_Z)$ value/order</td>
<td>0.1365/LO</td>
<td>0.130/LO</td>
</tr>
<tr>
<td>$a_s^{\text{ME}}(m_Z)$ value/order</td>
<td>0.130/LO</td>
<td>0.130/LO</td>
</tr>
<tr>
<td>MultipartonInteractions:pT0Ref [GeV]</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>MultipartonInteractions:ecmPow</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>MultipartonInteractions:coreRadius</td>
<td>0.54</td>
<td>0.38</td>
</tr>
<tr>
<td>MultipartonInteractions:coreFraction</td>
<td>0.68</td>
<td>0.33</td>
</tr>
<tr>
<td>ColorReconnection:range</td>
<td>2.63</td>
<td>2.32</td>
</tr>
<tr>
<td>$\chi^2$/dof</td>
<td>0.89</td>
<td>0.54</td>
</tr>
</tbody>
</table>

CMS Tunes with Pythia8 Color Reconnection and Herwig7 are ongoing.

- CP1: NNPDF3.1 LO ($\alpha_s = 0.130$)
- CP2: NNPDF3.1 LO ($\alpha_s = 0.130$)
- CP3: NNPDF3.1 NLO ($\alpha_s = 0.118$)
- CP4: NNPDF3.1 NLO ($\alpha_s = 0.118$)
- CP5: NNPDF3.1 NNLO ($\alpha_s = 0.118$)
UE in TTbar

- UE Measurement for the first time at a scale > 2m_t
- Measurement unfolded to particle level
- Good agreement of POWHEG+PYTHIA8 with CUETP8M2T4 in UE
Top Mass: b jet/ps/Color-reconnection

- **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

Top quark mass from ttbar fully hadronic (2016)

CR: MPI or QCD based, Gluon-Move
Rivet & Particle-Level Objects

- 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/t tagging automatically included, no distinction between light quarks and gluons (but \( g \rightarrow bb/cc \) possible)

- Recently implemented Rivet definitions in CMSSW module: `ParticleLevelProducer`, writing a Rivet plugin of the analysis becomes fast and simple.

---

**Overview: parton, particle and detector level**

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

Markus Seidel
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

<table>
<thead>
<tr>
<th>Event generator</th>
<th>Parton shower</th>
</tr>
</thead>
<tbody>
<tr>
<td>POWHEG (v2)</td>
<td>Pythia 8</td>
</tr>
<tr>
<td>POWHEG</td>
<td>Herwig++</td>
</tr>
<tr>
<td>MadGraph5_aMC@NLO</td>
<td>Pythia 8 with FxFx (t\bar{t} + 0, 1, 2j)</td>
</tr>
<tr>
<td>MadGraph5 (LO)</td>
<td>Pythia 8 with MLM (t\bar{t} + 0, 1, 2, 3j)</td>
</tr>
<tr>
<td>MadGraph5_aMC@NLO</td>
<td>Herwig++</td>
</tr>
</tbody>
</table>

$t\bar{t}$+heavy flavour events are extracted from inclusive $t\bar{t}$ sample
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($D^\pm \to K^\mp \pi^\pm \pi^\pm$) 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.**
Sizable differences between 4FS and 5FS at low pT. 5FS closer to data. Gluon splitting dominated at small ∆R_{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 $f$:
  $$N_{eff} = \frac{\left(\sum_i w_i\right)^2}{\sum w_i^2} = N(1 - 2f)^2$$
- For 35% negative weights (common at for high jet-multilplicity/ high pt)
  $\Rightarrow 9\%$ effective events compared to $w_i = 1$

$V(\text{DATA-others, CR}) \ast V(\text{MC, SR})/V(\text{MC, CR})$

Examples of CR $\rightarrow$ SR translation

<table>
<thead>
<tr>
<th>Final state</th>
<th>Background</th>
<th>Control sample</th>
<th>Features of extrapolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(1)-lepton + jets + MET</td>
<td>W+jets/ttbar with a “lost” lepton</td>
<td>1(2)-lepton</td>
<td>Correct lepton ID inefficiencies in MC to data, generally rely on MC to describe lepton acceptance</td>
</tr>
<tr>
<td>0-lepton + jets + MET</td>
<td>Z(\nu\nu)+jets</td>
<td>Z(\ell\ell)+jets or $\gamma$+jets</td>
<td>Hadronic recoil (MET excluding leptons or photon) in CR proxy for MET in SR, differences between Z and $\gamma$</td>
</tr>
</tbody>
</table>

[SUSY MT2](https://arxiv.org/abs/1705.04650)
• We are OFF by $\sim 5x$ on CPU power when considering Moore’s law
• **HL-LHC salvation will come from software improvements, not from hardware**

Elizabeth Sexton-Kennedy

**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
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!

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 Qiana 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 $\alpha_s$ is fixed to the value extracted from the $t\bar{t}$ events described previously.

<table>
<thead>
<tr>
<th></th>
<th>CUETP8M1</th>
<th>CUETP8M2T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tune</td>
<td>pp 14</td>
<td>pp 14</td>
</tr>
<tr>
<td>Tune</td>
<td>ee 7</td>
<td>ee 7</td>
</tr>
<tr>
<td>MultipartonInteractions</td>
<td>0.2521</td>
<td>0.2521</td>
</tr>
<tr>
<td>ecmPow</td>
<td>0.1365</td>
<td>0.1108</td>
</tr>
<tr>
<td>SpaceShower:alphaSvalue</td>
<td>NNPDF23.lo.qed.as.0130</td>
<td>NNPDF30.lo.as.0130</td>
</tr>
<tr>
<td>PDF pSet</td>
<td>LHAPDF6</td>
<td></td>
</tr>
<tr>
<td>MultipartonInteractions</td>
<td>2.40</td>
<td>2.20</td>
</tr>
<tr>
<td>spT0Ref</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>MultipartonInteractions</td>
<td>1.8</td>
<td>6.6</td>
</tr>
<tr>
<td>expPow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ColourReconnection:range</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CMS-PAS-TOP-16-021

Tune $\alpha_s^{ISR}$ using 8 TeV $t\bar{t}$ Njets and jet $p_T$ data $\rightarrow$

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

$\alpha_s^{ISR} = 0.1108^{+0.0145}_{-0.0142}$

$h_{c\text{omp}} = 1.581^{+0.658}_{-0.583} \times m_t$

$\Rightarrow$ Significantly lower shower $\alpha_s$ cures the overshoot of CUETP8M1 at high jet multiplicities.

$\Rightarrow$ UE and min-bias are described better

$\Rightarrow$ POWHEG+PYTHIA8: generally consistent with data, with residual differences covered by theory uncertainties.

Efe Yazgan

arXiv:1803.0399
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

$\alpha_s^{FSR}$ from jet substructure in ttbar $l+\text{jets}$ events

- Measured using charged+neutral and with only charged jet constituents (particle $p_T > 1 \text{ GeV}$).
- $b$, light, or gluon jet enriched samples.

$\text{PRD}.98.092014$

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