Measurement of jet substructure in $t\bar{t}$ events at 13 TeV

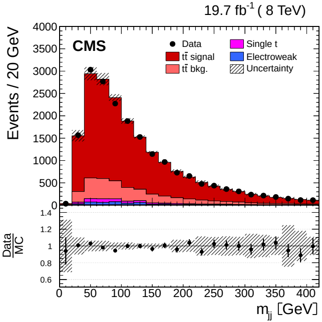
Markus Seidel

CERN

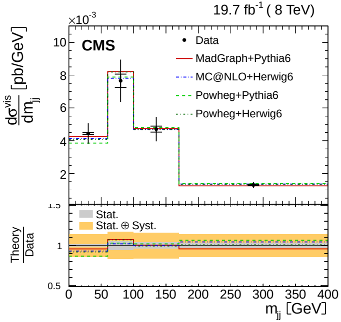
April 11, 2018

How experimentalists can contribute to generator development

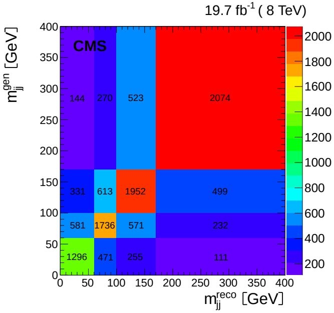
- Measure distributions that are sensitive to generator choices and parameters
- General problem: distributions at detector level can only be reproduced with running the whole detector simulation, and usually only by original authors
- Unfold to particle level so that theorists can easily compare with their latest models
    - Several tools on the market. Recommendation: TUnfold
- Implement in RIVET framework (>400 analyses from LEP, LHC and others)



=

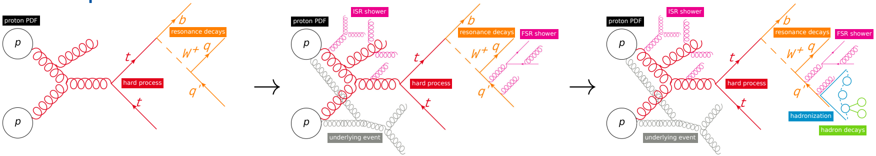


×



Define observables at particle level

Particle level



- No access to quarks and gluons, only hadrons, leptons and photons
- Not matching with current higher-order + resummation calculations for colored particles
- Exact match between different generators! Can be compared to future MC
- Usually close to detector level, using same algorithms for reconstructing resonances
- "*report only what you can see*" → small extrapolation uncertainties
- Use for solid comparison of data with MC generators

LHE level: Serious problems occur at NLO: Unphysical for MCatNLO method

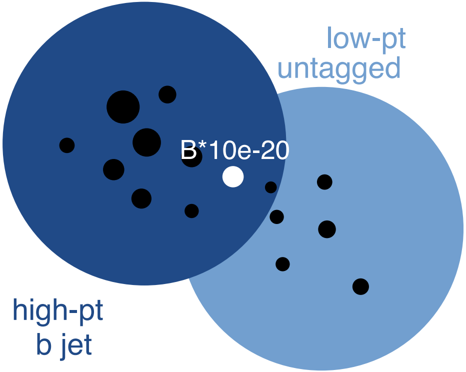
- Don't ever use LHE kinematics!

Parton level: Access to quarks and gluons without hadronization corrections

- Approximate match with higher-order + resummation calculations, but not exact
- *"The unfortunate truth is that most of the event record is intended for generator debugging rather than physics interpretation."* – Rivet manual

B jets at particle level

- Include B hadrons in jet clustering
    - Last hadron before the weak decay
- Four-momentum scaled by 10^{-20} → “ghost”
- Multi-tag is possible, e.g. $b\bar{b}$
- Similar for charm hadrons and tau mesons
- No parton level information
    - No distinction $t \to bW$ vs. $g \to b\bar{b}$
    - No distinction light quark vs. gluon



Jet substructure in $t\bar{t}$ events

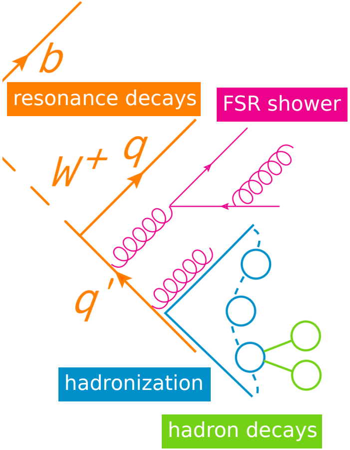
CMS PAS TOP-17-013

Motivation

- Fragmentation of quarks and gluons to jets described by parton shower + hadronization model
- Current models are tuned to LEP $Z \to q\bar{q}$ data
- Uncertainties relevant for many measurements, e.g. top mass

Measurement in $t\bar{t} \to$ lepton+jets

- "Standard candle" in pp collisions
- Jet substructure for each flavor: bottom, light-enriched, gluon-enriched
- Exhaustive analysis: more than 20 observables
  - Include recent propositions, useful for flavor discrimination + boosted heavy objects
  - Improve/tune tools for FSR shower and hadronization
- ATLAS already measured jet shapes in $t\bar{t}$ arXiv 1307.5749



Phase space definition

$t\bar{t} \rightarrow$ lepton+jets selection (follows closely reconstruction level selection)

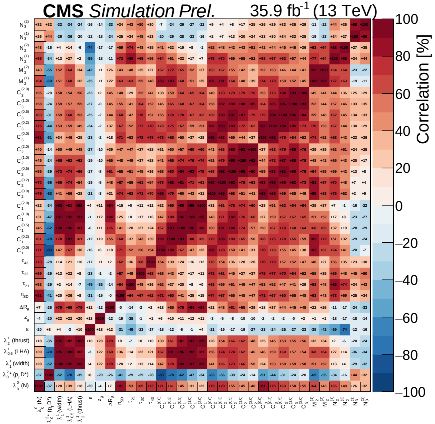
1. Exactly 1 electron or muon with $p_{\mathrm{T}} > 26\,\mathrm{GeV}, |\eta| < 2.4$,
   veto event in presence of additional lepton with $p_{\mathrm{T}} > 15\,\mathrm{GeV}, |\eta| < 2.4$
2. Require ≥ 4 jets with anti–k_T, R=0.4, $p_{\mathrm{T}} > 30\,\mathrm{GeV}, |\eta| < 2.5$
3. Require exactly 2 b-tagged jets
4. Require ≥ 2 untagged jets to form W candidate with $|m_{jj} - 80.4\,\mathrm{GeV}| < 15\,\mathrm{GeV}$

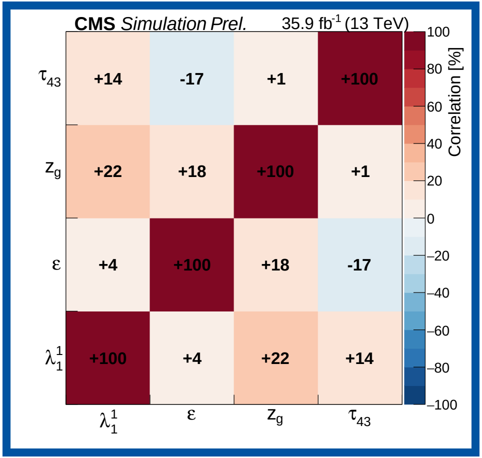
Jet flavor samples

- inclusive jets
- bottom jets: “ghost” B hadron clustered in jet (99% from b quarks)
- light-quark enriched jets: from W candidate (50% light quarks, 21% charm, 29% gluons)
- gluon-enriched jets: not from W candidate (1% bottom, 11% charm, 31% light, 58% gluon)

Overview of observables

- Generalized angularities λ_{β}^{κ}, eccentricity, soft drop observables, N-subjettiness, energy correlation function ratios
- Measured using charged, charged+neutral particles for different flavors → 264 distributions
- Large correlations → find set of 4 low-correlation observables

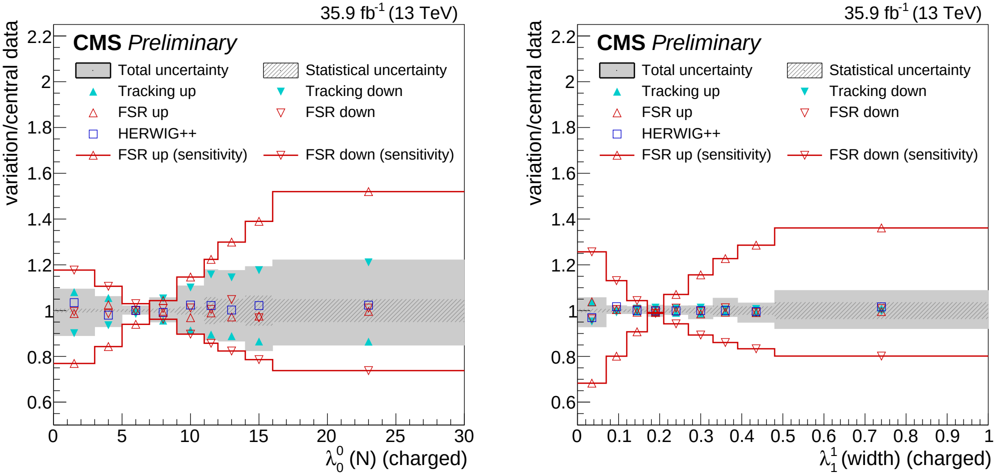




	λ_1^1	ε	z_g	τ_{43}
τ_{43}	+14	-17	+1	+100
z_g	+22	+18	+100	+1
ε	+4	+100	+18	-17
λ_1^1	+100	+4	+22	+14

Uncertainties

- Unfolding without regularization, verified with toy-experiments → no bias, correct σ_{stat}
- Systematic uncertainties on data/MC corrections + MC variations + bkg fractions, taken into account by unfolding data with alternative migration matrices
- Uncertainty on tracking efficiency (3–6%) can become dominant, depending on observable

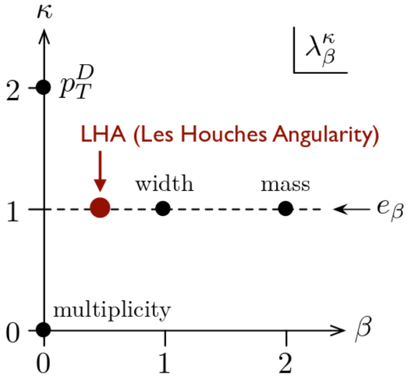


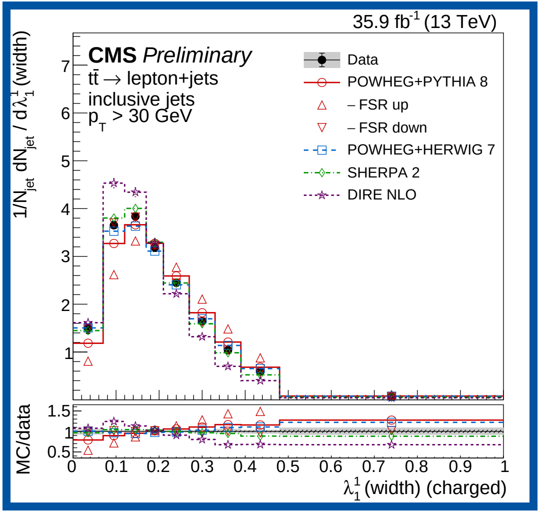
Generalized angularities: width

Defined in arXiv 1402.2657 as

$$\lambda^{\kappa}_{\beta} = \sum_i z_i^{\kappa} \left(\frac{\Delta R\left(i, \hat{n}_r\right)}{R} \right)^{\beta}$$

with $z_i = p_T^i / \sum_i p_T^i$ and recoil-free axis $\hat{n}_r$





- ⋆ Dire (NLO) 2.001: full $b \rightarrow bg$ structure not covered yet
- ○ Pythia 8 requires FSR down

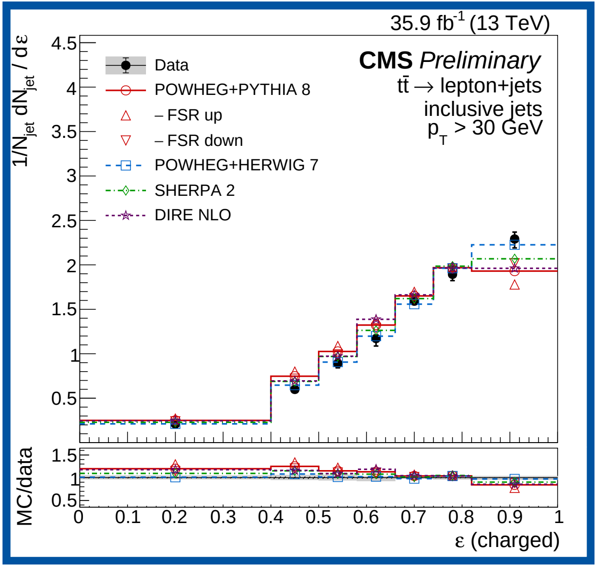
Eccentricity ε arXiv 1002.3982

- $\varepsilon = 1 - \frac{v_{\text{min}}}{v_{\text{max}}}$ with the eigenvalues of

$$M = \sum_i E_i \times \begin{pmatrix} (\Delta\eta_{i,\hat{n}_r})^2 & \Delta\eta_{i,\hat{n}_r}\Delta\phi_{i,\hat{n}_r} \\ \Delta\phi_{i,\hat{n}_r}\Delta\eta_{i,\hat{n}_r} & (\Delta\phi_{i,\hat{n}_r})^2 \end{pmatrix}$$

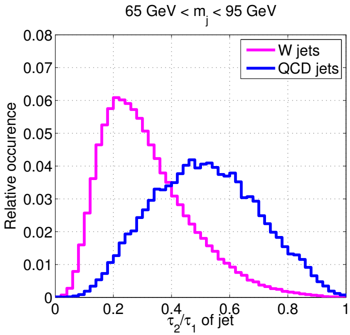
- Perfectly circular jet: $\varepsilon = 0$
- Elliptical jet: $\varepsilon \to 1$

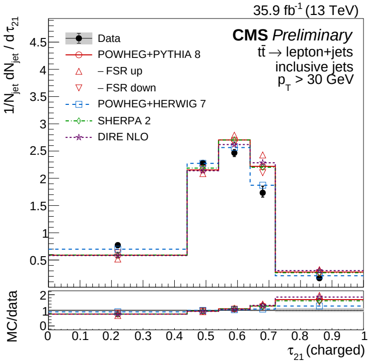
- Best agreement with ☐ Herwig 7

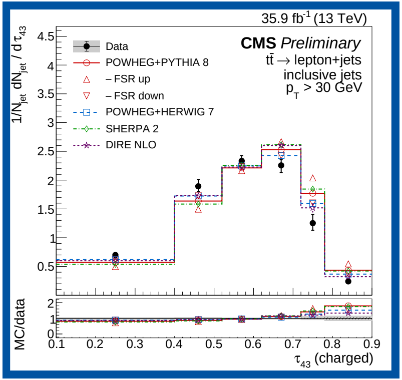


N-subjettiness ratios arXiv 1108.2701

- $\tau_{NM} = \tau_N / \tau_M$, where $\tau_\mathrm{N}^{(\beta)} = \frac{1}{d_0} \sum_i p_{\mathrm{T},i} \min \left\{ (\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta, \ldots, (\Delta R_{N,i})^\beta \right\}, \beta = 1$
  - + minimization step for finding subjet axes that minimize τ_N
- Used for distinguishing jets with N or M subjets, e.g. τ_{21} in W tagging



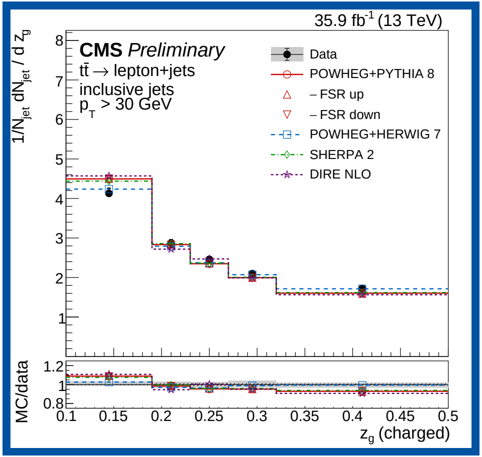


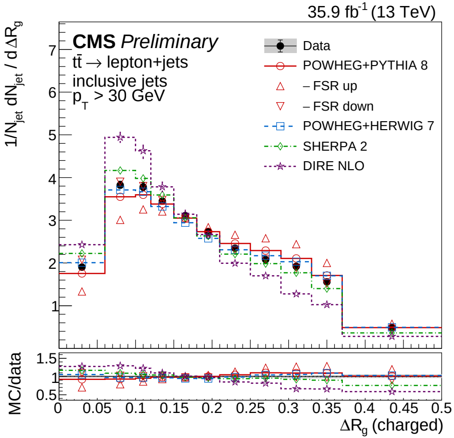


- Data not well described by MC generators (correlated with charged multiplicity)

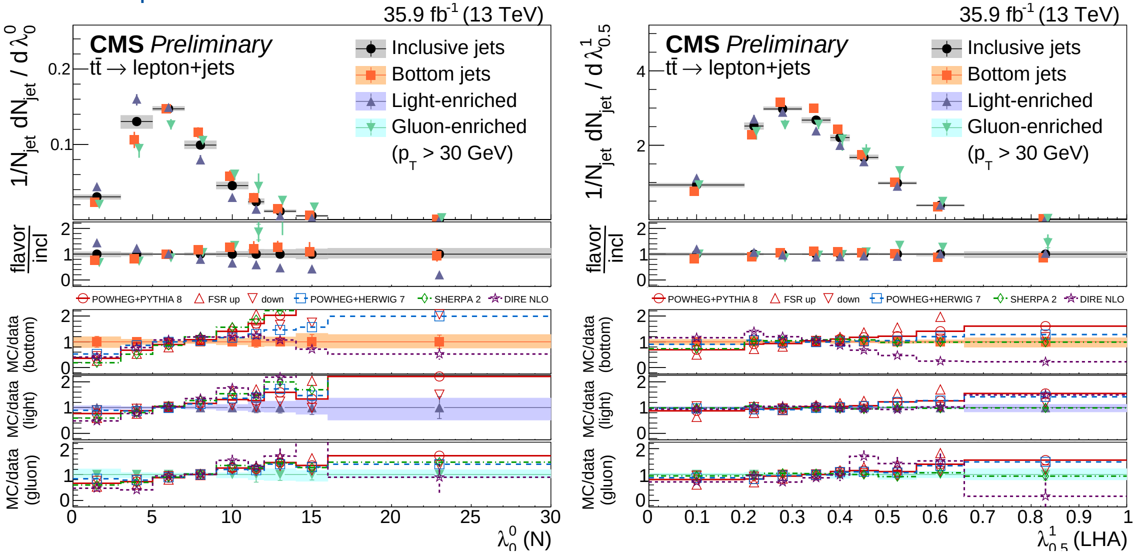
Soft drop observables arXiv 1704.05066

- Iterative soft drop declustering $j_0 \to j_1 + j_2, j_1 \to j_0$, stop when $z_g = p_\mathrm{T}(j_2)/p_\mathrm{T}(j_0) > 0.1$
- Groomed momentum fraction z_g of last iteration
  - Related to QCD splitting function, independent of $\alpha_s^\mathrm{FSR}(m_Z)$
  - Best agreement with ☐ Herwig 7 (angular-ordered)
- ΔR_g = angle between subjets, related to jet width and groomed area





Flavor dependence



- Left: MCs do not describe charged multiplicity (λ_0^0), bottom jets seem worse
- Right: Les Houches Angularity described equally well for all flavors

χ^2 tests

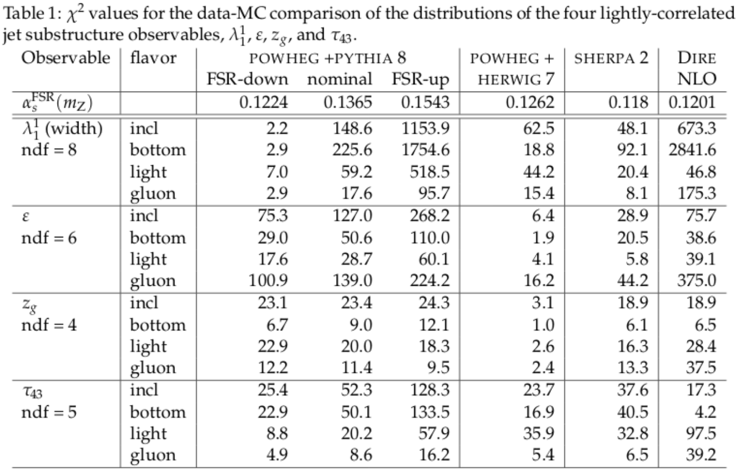


Table 1: χ^2 values for the data-MC comparison of the distributions of the four lightly-correlated jet substructure observables, λ_1^1, ε, z_g, and τ_{43}.

Observable	flavor	POWHEG +PYTHIA 8			POWHEG +	SHERPA 2	DIRE
		FSR-down	nominal	FSR-up	HERWIG 7		NLO
$\alpha_s^{\mathrm{FSR}}(m_Z)$		0.1224	0.1365	0.1543	0.1262	0.118	0.1201
λ_1^1 (width)	incl	2.2	148.6	1153.9	62.5	48.1	673.3
ndf = 8	bottom	2.9	225.6	1754.6	18.8	92.1	2841.6
	light	7.0	59.2	518.5	44.2	20.4	46.8
	gluon	2.9	17.6	95.7	15.4	8.1	175.3
ε	incl	75.3	127.0	268.2	6.4	28.9	75.7
ndf = 6	bottom	29.0	50.6	110.0	1.9	20.5	38.6
	light	17.6	28.7	60.1	4.1	5.8	39.1
	gluon	100.9	139.0	224.2	16.2	44.2	375.0
z_g	incl	23.1	23.4	24.3	3.1	18.9	18.9
ndf = 4	bottom	6.7	9.0	12.1	1.0	6.1	6.5
	light	22.9	20.0	18.3	2.6	16.3	28.4
	gluon	12.2	11.4	9.5	2.4	13.3	37.5
τ_{43}	incl	25.4	52.3	128.3	23.7	37.6	17.3
ndf = 5	bottom	22.9	50.1	133.5	16.9	40.5	4.2
	light	8.8	20.2	57.9	35.9	32.8	97.5
	gluon	4.9	8.6	16.2	5.4	6.5	39.2

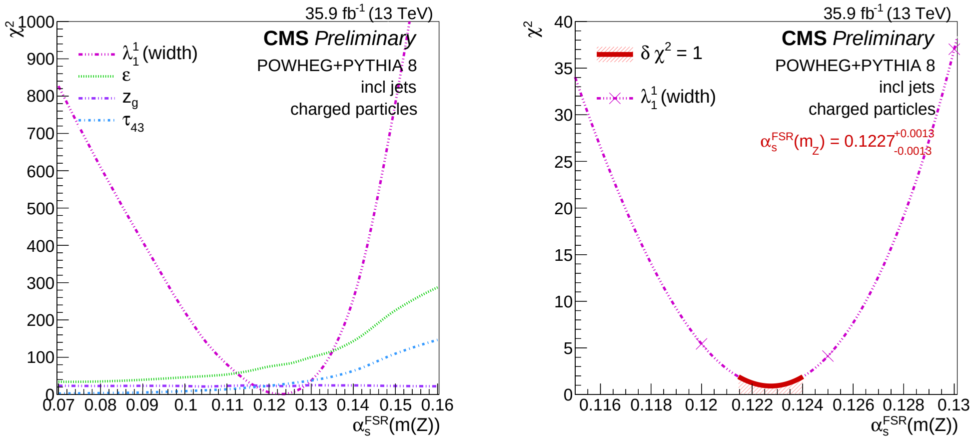
Main setup comparison

- Pythia FSR down is preferred over default
- Herwig 7 agrees very well for ε and z_g

Other Pythia settings

- b jets: prefer harder b fragmentation
- light quarks / gluons: prefer QCD-based color reconnection

Scan of Powheg+Pythia 8 $\alpha_s^{\mathrm{FSR}}(m_Z)$



- Need more complete tuning to get agreement with all observables
- From jet width only → $\alpha_s^{\mathrm{FSR}}(m_Z) = 0.1227 \pm 0.0013 \approx$ FSR down
- Comparison to world average would require conversion to CMW scheme and scale uncertainties ($\rightsquigarrow {}^{+0.014}_{-0.012}$)

Summary

- Measured various jet substructure observables probing the jet evolution
  - Angularities, N-subjettiness, soft drop, energy correlations...
  - All observables available with charged and charged+neutral particles, for inclusive, bottom, light-enriched and gluon-enriched jet samples
  - Too many to show in paper – full results will be available in HepData and Rivet

- Comparisons with recent generators: Powheg+Pythia 8/Herwig 7, Sherpa, Dire NLO
  - No generator can describe all observables → tuning needed
  - Improvement of Dire NLO shower expected: cover full structure of $b \to bg$ branchings in next version. Promises reduced uncertainties wrt other parton showers

- Demonstrated simple tuning of Powheg+Pythia 8 $\to \alpha_s^{\mathrm{FSR}}(m_Z) = 0.1227 \pm 0.0013$ → validation of FSR scale uncertainty range (factor of 2 for shower μ_R)
- Constraints on color reconnection and hadronization models possible