Multi-Jet Production and other DIS developments in Pythia

Joni Laulainen[1] C.T. Preuss[2] I. Helenius[1]

[1]University of Jyväskylä
Department of Physics

[2]University of Wuppertal
Department of Physics

[hep-ph/2410.20950]
Accepted to JHEP

Pythia Week, May 14th 2025

Multi-Jet Production and other DIS developments in Pythia
Joni Laulainen
Jet production in DIS
Parton Showers
Motivation
Merging
Merging scale
Results
HERA data comparisons
Scale variations
Tuning
DIS NLOPS

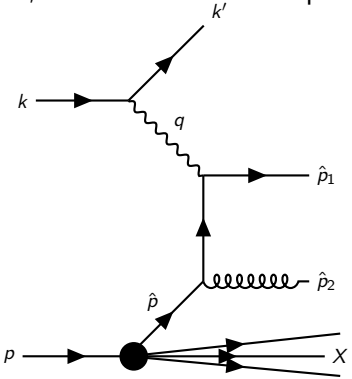
Jet production in DIS

- DIS modeling lacking behind LHC-driven improvements in PYTHIA.
- Good description of DIS jets important with EIC in mind.
- Multi-jet production has multiple hard scales, hardest scale not unique.

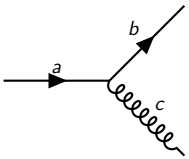
$$\mu_F^2 = Q^2 \qquad \mu_F^2 = \left(Q^2 + H_T^2\right)/2$$

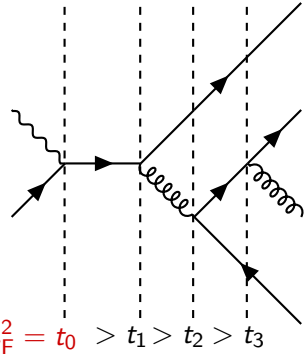
$$\text{where } H_T = \sum_{i=1}^{N} \hat{p}_{T,i}$$

seen success in [J. Currie et al., JHEP 07 (2017) 018] and [S. Höche et al., Phys.Rev.D 98 (2018) 11]



Parton Showers



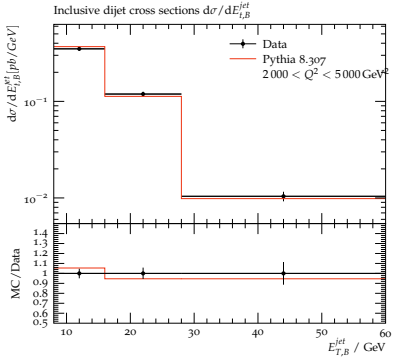


When producing events in PYTHIA, the shower evolution usually goes from μ_F^2 to Λ_{QCD}. In DIS, the recommended option is to use the **dipole recoil**[1] shower with the **power shower** option ($t_0 \approx \sqrt{s}$).

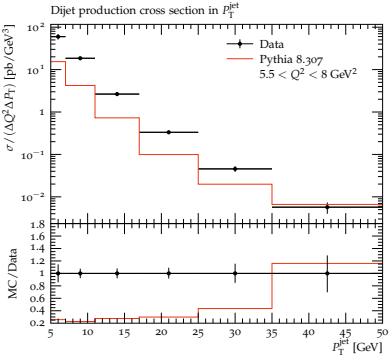
[1] [B. Cabouat, T. Sjöstrand, Eur.Phys.J.C 78 (2018) 3, 226]

Motivation - Parton shower approximation for dijets

- Good agreement in high-Q^2 using power showers (dipole recoil)
- Poor description of low-Q^2 dijets



[ZEUS Collaboration, Eur.Phys.J.C 70 (2010) 965-982]

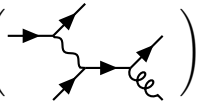


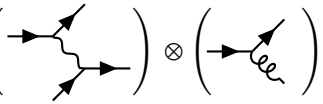
[H1 Collaboration, Eur.Phys.J.C 77 (2017) 4, 215]

Multi-Jet Merging

- Objective: Combine parton showers with matrix elements
- Separate phase space using merging scale.

$$\text{Hard jets} : p_{\perp}^2 > t_{\text{MS}} \qquad \text{Soft jets} : p_{\perp}^2 \leq t_{\text{MS}}$$

(ME) 

(PS) 

SHERPA generated parton level events (ME) up to +4 jets. MADGRAPH5 reliably up to +2 jets.
[T. Gleisberg et al., JHEP 02 (2009) 007]
[J. Alwall et al., JHEP 07 (2014) 079]

VINCIA as the parton shower (PS) with its sector merging machinery
[C.T.Preuss et al., JHEP 07 (2020) 032]

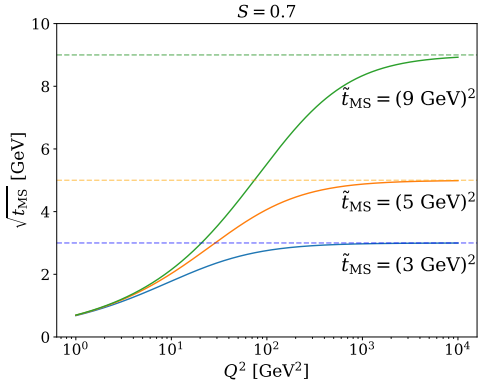
Merging - t_{MS}-dependence

Phase space separation with dynamic merging scale[2].

- Attempt to improve low-Q^2 jet production.
- Let MEs contribute more in low-Q^2 region.

$$t_{\mathrm{MS}} = \frac{\tilde{t}_{\mathrm{MS}}}{1 + \frac{\tilde{t}_{\mathrm{MS}}}{S^2 Q^2}}$$

$$\tilde{t}_{\mathrm{MS}} = 5\ \mathrm{GeV}, S = 0.7$$



[2] SHERPA DIS jet merging [T. Carli, T. Gehrmann, S. Höche, Eur.Phys.J.C 67 (2010) 73-97]

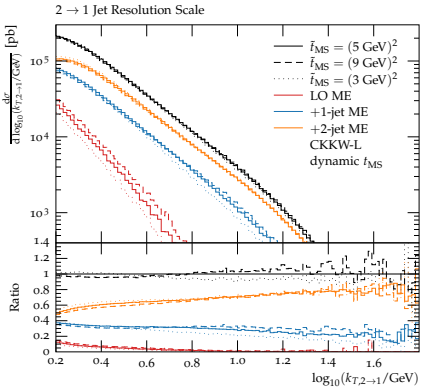
Merging - t_{MS}-dependence

Phase space separation with dynamic merging scale[2].

- Attempt to improve low-Q^2 jet production.
- Let MEs contribute more in low-Q^2 region.

$$t_{\mathrm{MS}} = \frac{\tilde{t}_{\mathrm{MS}}}{1 + \frac{\tilde{t}_{\mathrm{MS}}}{S^2 Q^2}}$$

$$\tilde{t}_{\mathrm{MS}} = 5\ \mathrm{GeV}, S = 0.7$$

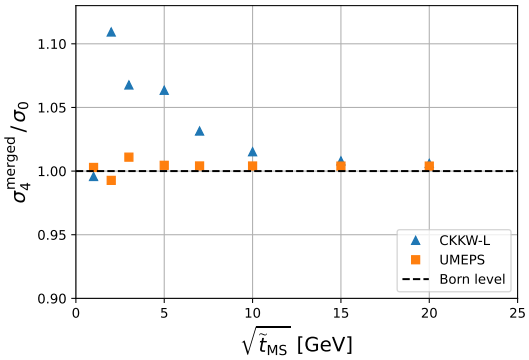


[2] SHERPA DIS jet merging [T. Carli, T. Gehrmann, S. Höche, Eur.Phys.J.C 67 (2010) 73-97]

Merging - Algorithms

Merging algorithm effect on inclusive cross section. DIS +4 jets.

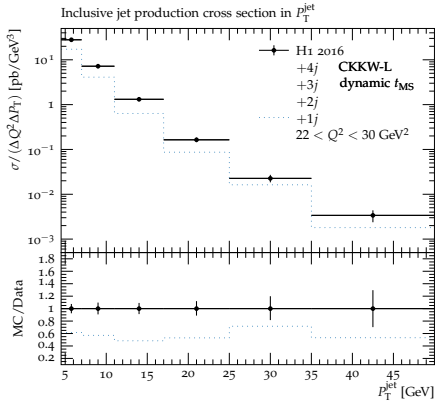
- CKKW-L[1], dynamic merging
  - phase space separation t_{MS} and reweighting using PS histories to account for double-counting.
- UMEPS[2], fixed merging
  - similar, but preserves unitarity through "add-and-subtract" scheme.

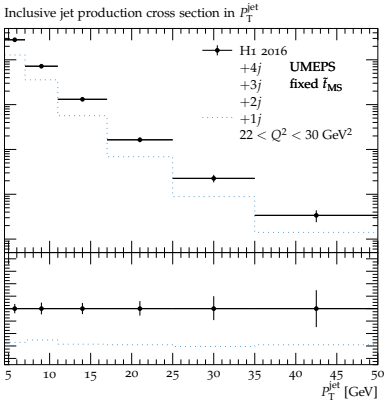


[1] [L. Lönnblad, JHEP 05 (2002) 046]
[2] [L. Lönnblad, S. Prestel, JHEP 02 (2013) 094]

Results - Inclusive jets

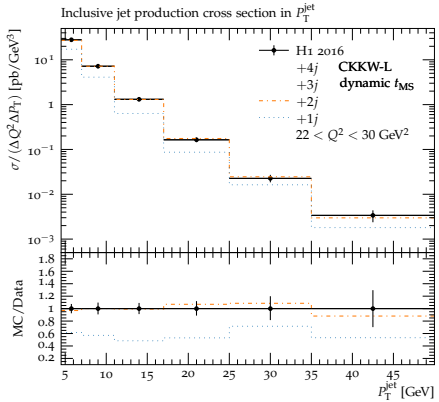
We compare the results to HERA data. [H1 Collaboration, Eur.Phys.J.C 77 (2017) 4, 215]

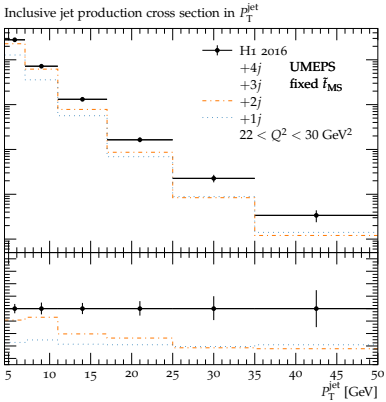




Results - Inclusive jets

We compare the results to HERA data. [H1 Collaboration, Eur.Phys.J.C 77 (2017) 4, 215]

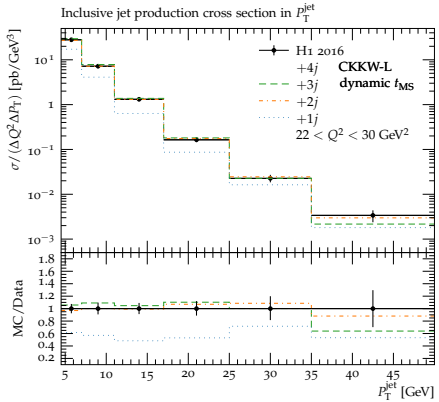


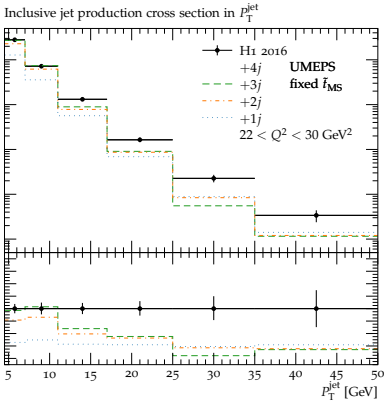


Satisfactory result with just 2 additional jets.

Results - Inclusive jets

We compare the results to HERA data. [H1 Collaboration, Eur.Phys.J.C 77 (2017) 4, 215]

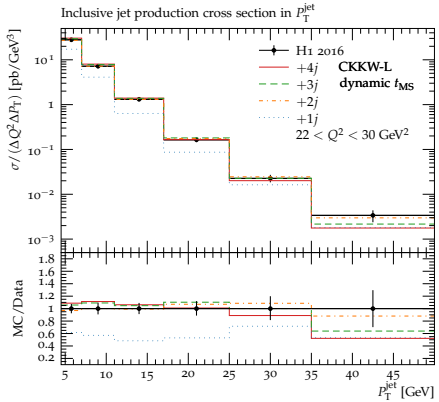


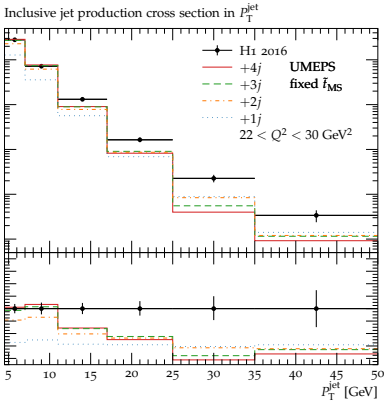


Satisfactory result with just 2 additional jets.

Results - Inclusive jets

We compare the results to HERA data. [H1 Collaboration, Eur.Phys.J.C 77 (2017) 4, 215]



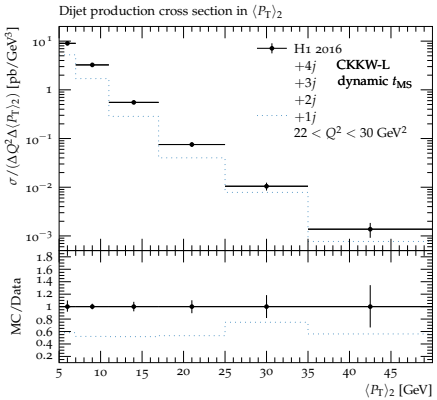


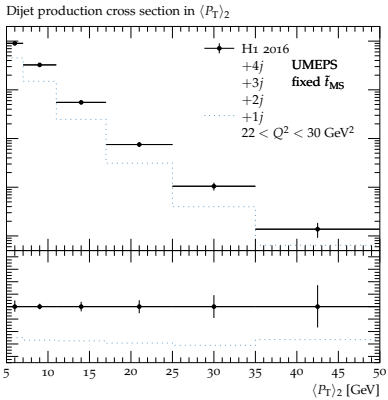
Satisfactory result with just 2 additional jets.

Result converges with addition of higher multiplicity samples.

Results - Dijets

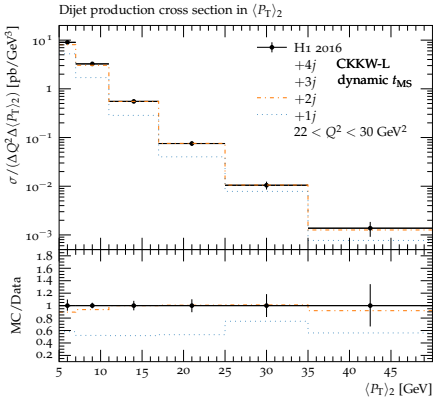
Similar outcome as with inclusive jets. UMEPS fails to match data at high-p_T.

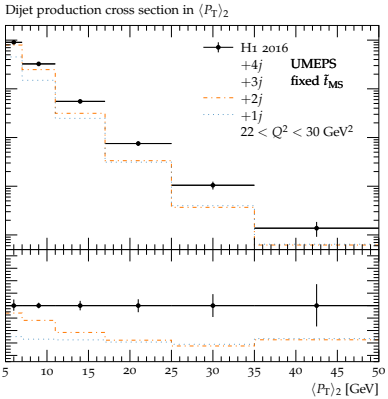




Results - Dijets

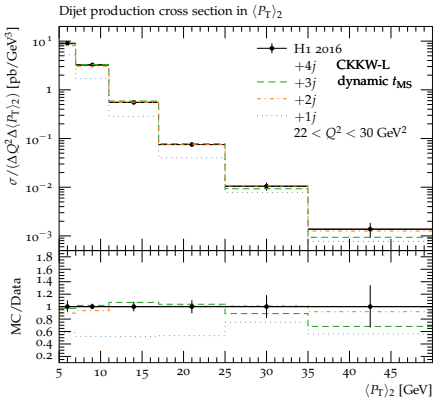
Similar outcome as with inclusive jets. UMEPS fails to match data at high-p_T.

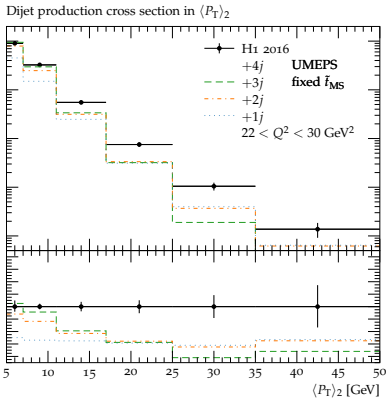




Results - Dijets

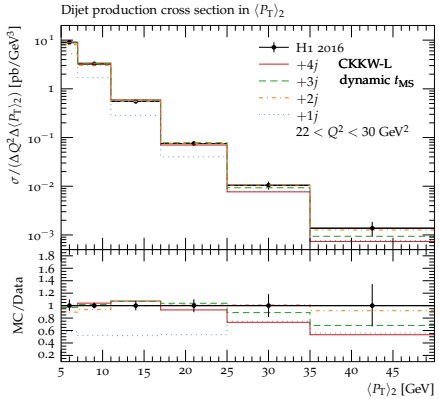
Similar outcome as with inclusive jets. UMEPS fails to match data at high-p_T.

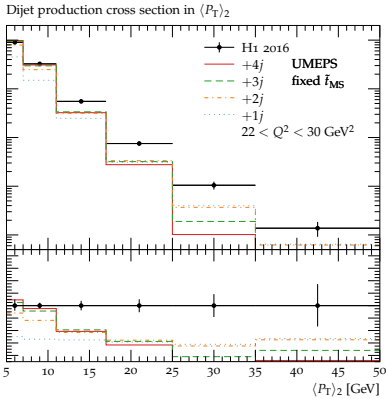




Results - Dijets

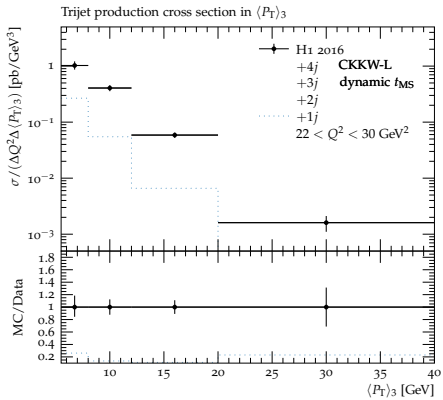
Similar outcome as with inclusive jets. UMEPS fails to match data at high-p_T.

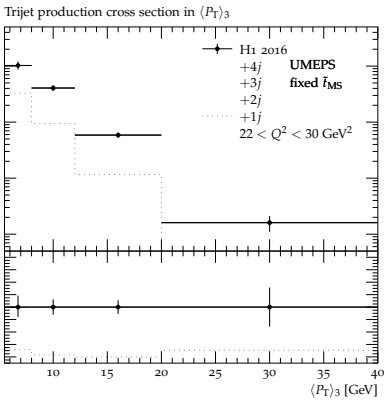




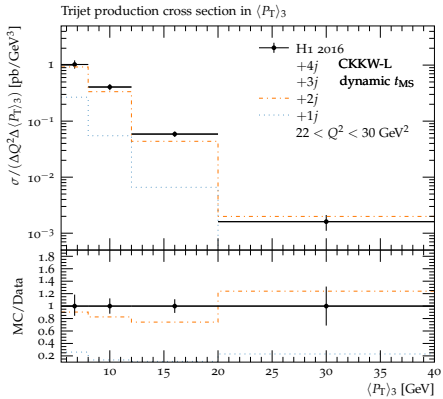
[Eur.Phys.J.C 77 (2017) 4, 215] RIVET analysis released (along with [Phys.Lett.B 542 (2002) 193-206]).

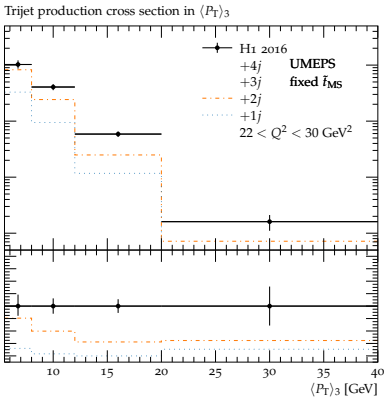
Results -Trijets



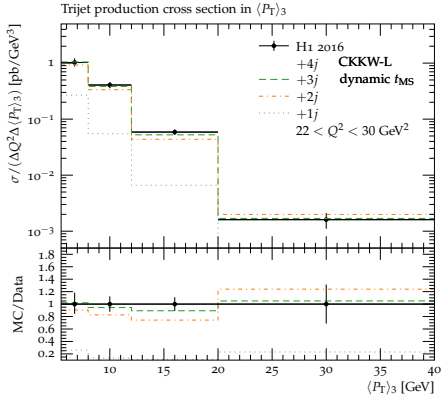


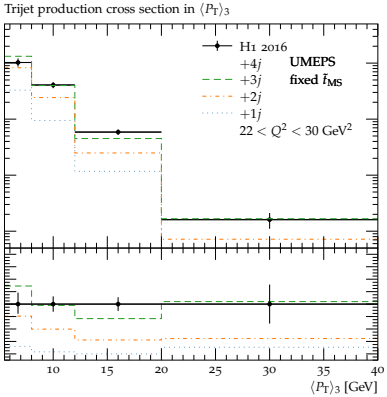
Results -Trijets



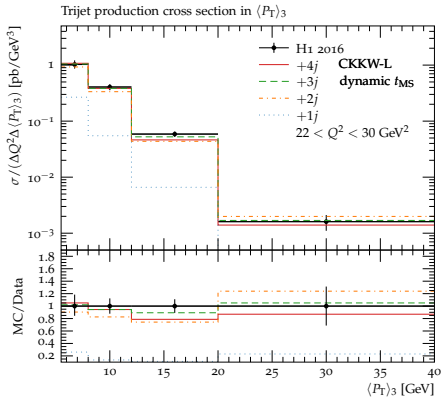


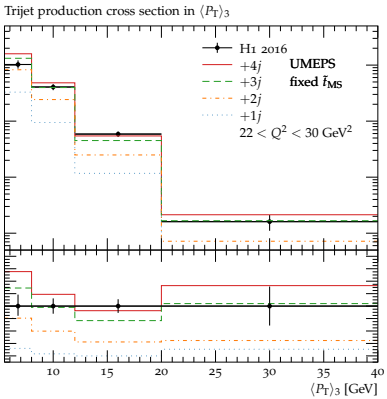
Results -Trijets





Results -Trijets

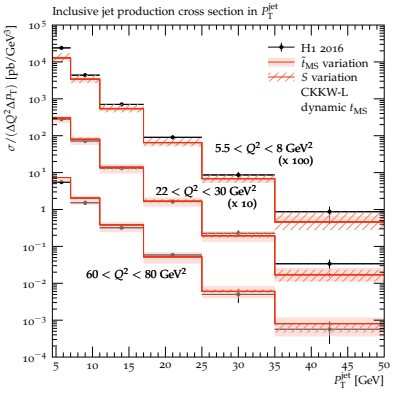


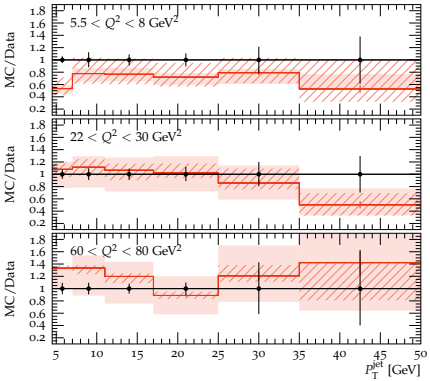


Competent trijet modeling requires at least +3-jet merging.

Results - CKKW-L parameter variations

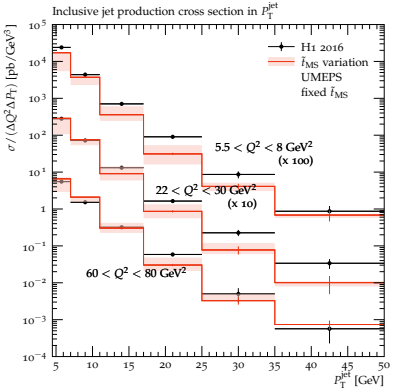
Variation of $S = 0.7 \pm 0.1$ and $\sqrt{\tilde{t}_{\mathrm{MS}}} = 5$ between 3 and 9 GeV.

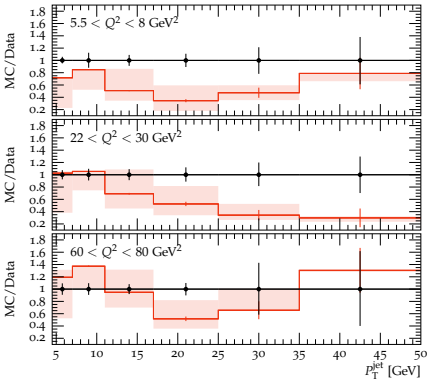




S-parameter mostly influences low-Q^2 region.

Results - UMEPS parameter variations

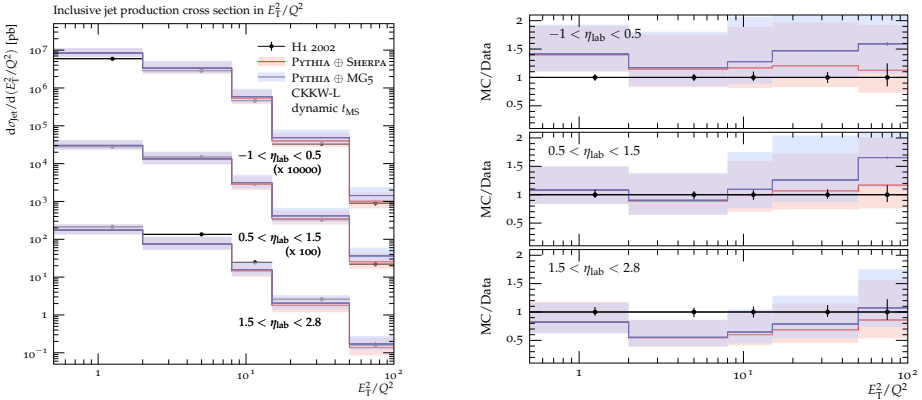




Moderate sensitivity to fixed merging scale variations.

Results - Scale variations

Also possible to use MG5 input. CKKW-L +2 jets.
Uncertainty bands by varying the scales μ_R and μ_F by a factor of 2.

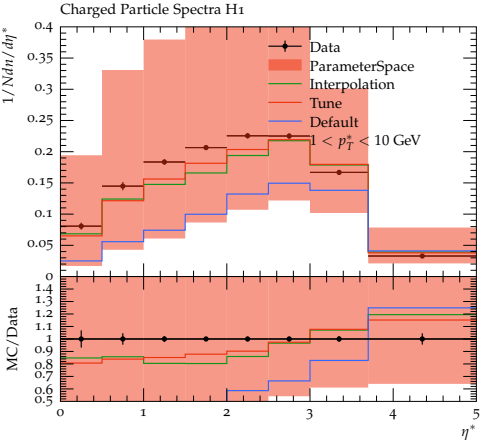


[H1 Collaboration, Phys.Lett.B 542 (2002) 193-206]

Tuning in progress

- In DIS, need to use dipole recoil option for default shower
- Hadronization parameters have not been tuned with this option
- Use data from HERA and tune[1] using the Professor 2 tool

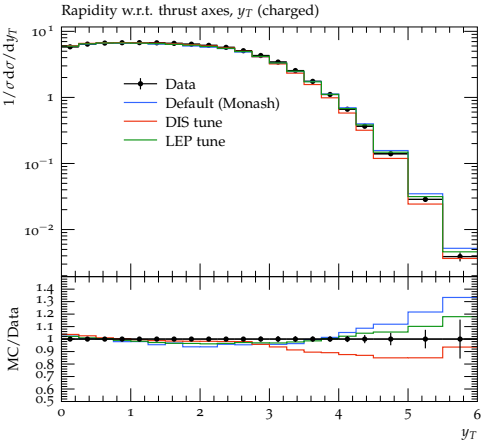
Parameter	Tune	Default	Min	Max
StringZ:aLund	1.24	0.68	0.40	1.60
StringZ:bLund	0.73	0.98	0.40	1.60
StringPT:sigma	0.474	0.335	0.250	0.800



[1] This is a very preliminary tune

Tuning in progress

- The same tune with LEP data shows signs of non-universality of string model parameters
- for HERA tune: $\chi^2/N_{\text{dof}} = 6.564$ with $(a = 1.24, b = 0.73, \sigma_{\text{pT}} = 0.474)$
- for LEP tune: $\chi^2/N_{\text{dof}} = 1.897$ with $(a = 0.45, b = 0.42, \sigma_{\text{pT}} = 0.208)$



DIS NLOPS in progress

- Matching first emission of default parton shower to NLO matrix element
- NLO prediction of observable O:

$$\langle O \rangle_{\text{NLO}} = \int \bar{B}\, O(\Phi_B)\, \mathrm{d}\Phi_B + \int R(\Phi_R)\, [O(\Phi_R) - O(\Phi_B)]\, \mathrm{d}\Phi_R,$$

$$\text{where } \bar{B} = \left[B(\Phi_B) + V(\Phi_B) + \int R(\Phi_R)\, \mathrm{d}\Phi_{+1} \right]$$

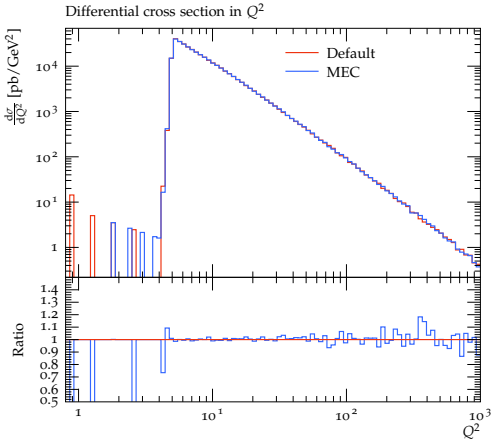
- Parton shower prediction to ($\mathcal{O}(\alpha_s)$):

$$\langle O \rangle^{(1)}_{\text{PS}} = \int \underbrace{\bar{B} \left[(\Phi_B)\, O(\Phi_B) + \int_{t_c}^{t_0} K_{B \mapsto R}(\Phi_R)\, [O(\Phi_R) - O(\Phi_B)]\, \mathrm{d}\Phi_{+1} \right]}_{\mathcal{F}_0^{(1)}(\Phi_B, O)} \mathrm{d}\Phi_B$$

DIS NLOPS in progress

- Matrix element corrections put in place in branch `jl-dis-nlops`

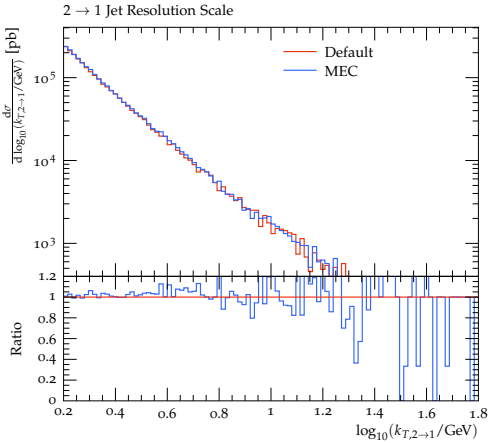
$$R_{\gamma^* q \to qg}(x, z_1) = \frac{\mathrm{d}\sigma^{\mathrm{ME}}_{\gamma^* q \to qg}/\mathrm{d}x\mathrm{d}z_1}{\mathrm{d}\sigma^{\mathrm{PS}}_{\gamma^* q \to qg}/\mathrm{d}x\mathrm{d}z_1}$$
$$= \frac{x^2 + z_1^2 + 2(1 + 3xz_1)(1 - x)(1 - z_1)}{1 + x^2}$$
$$R_{\gamma^* g \to q\bar{q}}(x, z_1)$$
$$= \frac{z_1^2 + (1 - z_1)^2}{z_1} + 8x\frac{(1 - x)(1 - z_1)}{x^2 + (1 - x)^2}$$



DIS NLOPS in progress

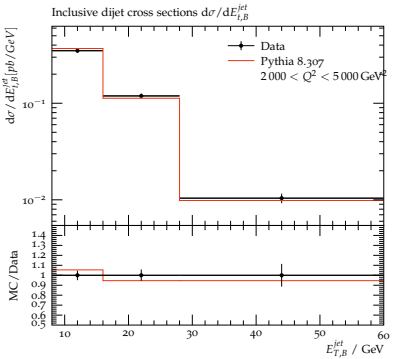
- Matrix element corrections put in place in branch `jl-dis-nlops`

$$R_{\gamma^* q \to qg}(x, z_1) = \frac{\mathrm{d}\sigma^{\mathrm{ME}}_{\gamma^* q \to qg}/\mathrm{d}x\mathrm{d}z_1}{\mathrm{d}\sigma^{\mathrm{PS}}_{\gamma^* q \to qg}/\mathrm{d}x\mathrm{d}z_1}$$
$$= \frac{x^2 + z_1^2 + 2(1 + 3xz_1)(1 - x)(1 - z_1)}{1 + x^2}$$
$$R_{\gamma^* g \to q\bar{q}}(x, z_1)$$
$$= \frac{z_1^2 + (1 - z_1)^2}{z_1} + 8x\frac{(1 - x)(1 - z_1)}{x^2 + (1 - x)^2}$$



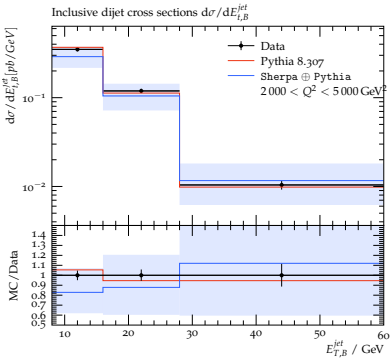
Summary

- Default PYTHIA shower not enough to describe jets in DIS
  $\rightarrow$ merging implementation.
- Problem: Hardest scale not unique
  $\rightarrow$ use dynamic merging scale or different factorization scale choice.
- Recommended setup: CKKW-L with dynamic t_{MS} and $\mu_{\mathrm{F}} = Q^2$
  $\rightarrow$ good description of HERA data also in low-Q^2 region.
- Also tuning efforts with HERA data and a NLO-matching project underway.



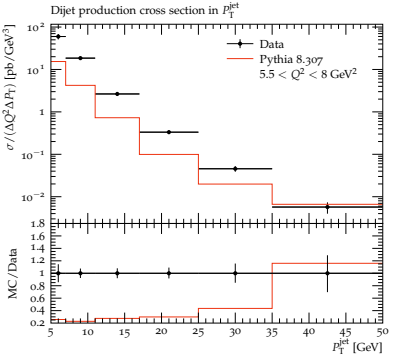
Summary

- Default PYTHIA shower not enough to describe jets in DIS
  $\rightarrow$ merging implementation.
- Problem: Hardest scale not unique
  $\rightarrow$ use dynamic merging scale or different factorization scale choice.
- Recommended setup: CKKW-L with dynamic t_{MS} and $\mu_{\mathrm{F}} = Q^2$
  $\rightarrow$ good description of HERA data also in low-Q^2 region.
- Also tuning efforts with HERA data and a NLO-matching project underway.



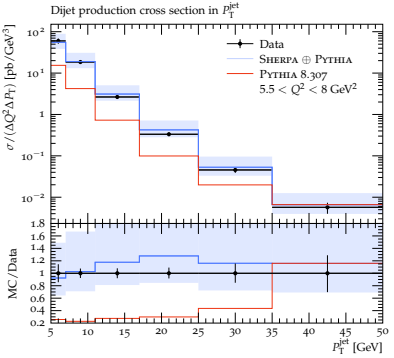
Summary

- Default PYTHIA shower not enough to describe jets in DIS
  $\rightarrow$ merging implementation.
- Problem: Hardest scale not unique
  $\rightarrow$ use dynamic merging scale or different factorization scale choice.
- Recommended setup: CKKW-L with dynamic t_{MS} and $\mu_{\mathrm{F}} = Q^2$
  $\rightarrow$ good description of HERA data also in low-Q^2 region.
- Also tuning efforts with HERA data and a NLO-matching project underway.

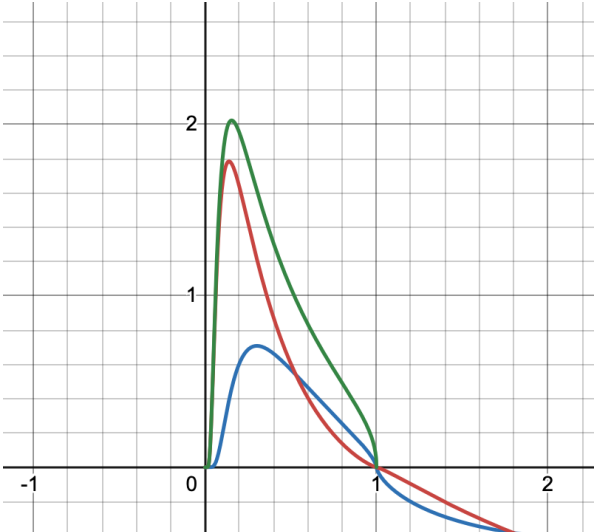


Summary

- Default PYTHIA shower not enough to describe jets in DIS
  $\rightarrow$ merging implementation.
- Problem: Hardest scale not unique
  $\rightarrow$ use dynamic merging scale or different factorization scale choice.
- Recommended setup: CKKW-L with dynamic t_{MS} and $\mu_{\mathrm{F}} = Q^2$
  $\rightarrow$ good description of HERA data also in low-Q^2 region.
- Also tuning efforts with HERA data and a NLO-matching project underway.



Lund fragmentation function



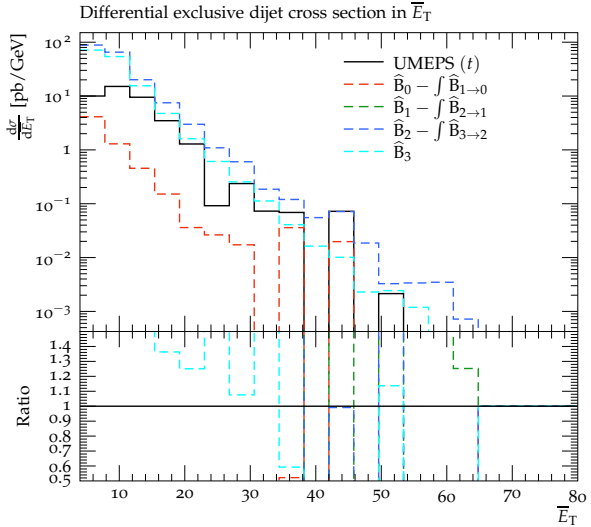
Backup slides: Fixed order calculations / Matrix elements

Why not just use exact fixed order calculations all the way to a given final state multiplicity? Why stop at $+4$ jets?
Number of diagrams grows almost **factorially** as a function of N.
Final states are observed to contain **hundreds** of particles.

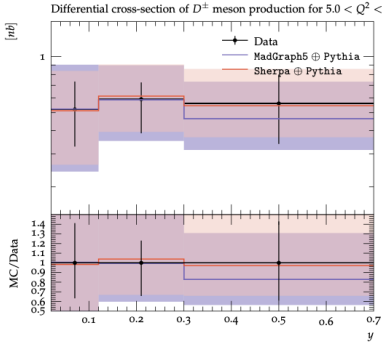
Process	# of diagrams
$e^- p \to e^- j$	20
$e^- p \to e^- 2j$	60
$e^- p \to e^- 3j$	720
$e^- p \to e^- 4j$	5 340
$e^- p \to e^- 5j$	71 080

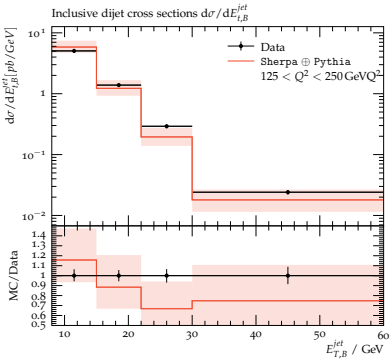
Access description of up to 5 hard jets with matrix elements.

UMEPS dynamical merging scale



Other analyses: ZEUS_2008_I810112 and ZEUS_2010_I875006





Hard events are generated at leading order in QCD using a HPC-enabled version of the Sherpa event generator[3] and stored using the efficient and scalable HDF5 event-file format. We use the G_μ-scheme with electroweak input parameters

$$m_W = 80.419\ \text{GeV}\,, \quad m_Z = 91.188\ \text{GeV}\,, \quad G_\text{F} = 1.16639 \times 10^{-5}\ \text{GeV}^{-2}\,.$$

Quark masses are taken in the four-flavour scheme and the b-quark mass is set to $m_b = 4.7$ GeV. The PDF choice is `NNPDF40_lo_pch_as_01180`.

[3] Available at `https://gitlab.com/hpcgen/me`.

I. Produce *Les Houches* event files (LHEF) [25] with a matrix element generator for $n = 0, 1 \ldots N$ extra jets with a regularisation cut-off, ρ_{MS}, typically using a fixed factorisation scale, μ_F, and a fixed $\alpha_s(\mu_R)$.

II. Pick a jet multiplicity, n, and a state S_{+n} according to the cross sections given by the matrix element generator.

1. Find all shower histories for the state S_{+n}, pick a sequence according to the product of splitting probabilities. Only pick un-ordered sequences if no ordered sequence was found. Only pick incomplete paths if no complete path was constructed.
2. Perform reweighting: For each $0 \leqslant i < n$,
   i. Start the shower off the state S_{+i} at ρ_i, generate a trial state R with scale ρ_R. If $\rho_R > \rho_{i+1}$, veto the event and start again from II.
   ii. Calculate the weight factor

$$w_i = \frac{\alpha_s(\rho_{i+1})}{\alpha_s(\mu_R)} \frac{x_i^+ f_i^+(x_i^+, \rho_i)}{x_i^+ f_i^+(x_i^+, \rho_{i+1})} \frac{x_i^- f_i^-(x_i^-, \rho_i)}{x_i^- f_i^-(x_i^-, \rho_{i+1})} \tag{4.16}$$

3. Start the shower from S_{+n}.
   i. If $n < N$, start the shower at ρ_n, veto any shower emission producing an additional resolved jet.
   ii. If $n = N$, start the shower at ρ_n.

III. If the event was not rejected, multiply the event weight by

$$w'_n = \frac{x_n^+ f_n^+(x_n^+, \rho_n)}{x_n^+ f_n^+(x_n^+, \mu_F)} \times \frac{x_n^- f_n^-(x_n^-, \rho_n)}{x_n^- f_n^-(x_n^-, \mu_F)} \times \prod_{i=0}^{n-1} w_i \tag{4.17}$$

V. Start again from II.

The second part, i.e. producing $\int_3 \widehat{\mathrm{B}}_{n \to m}$–events to effect lower-multiplicity PS resummation, requires only two changes:

II.3 Replace the matrix-element state by S_{+n-1}, or the first state S_{+l} with all $l \leq n-1$ partons above the merging scale. If no integrated state can be constructed, i.e. if only incomplete paths were found, reject the event. For valid events, start the shower at ρ_n, veto any shower emission producing an additional resolved jet.

1. Calculate cross sections and generate events according to step 1 and 2 in section 2.1. Q_{MS} denotes the merging scale which is equal to the matrix element cutoff and may be defined using any choice of scale. The events are generated using a fixed strong coupling, $\alpha_{s\text{ME}}$, and a maximum parton multiplicity, N.
2. Construct a full cascade history by considering all possible ordered histories and selecting one randomly with a probability proportional to the product of the branching probabilities. If no ordered histories can be constructed, unordered ones are considered. This results in a set of intermediate states $(S_2, S_3, \ldots S_n)$ and scales $(\rho_2 = \rho_{\max}, \rho_3, \ldots, \rho_n)$. S_2 denotes here denotes the constructed $2 \to 2$ process and S_n is the state given by the matrix element. n is the parton multiplicity in the event, $\rho_{\max}$ is the maximum scale of the process and ρ_i is the constructed scale where the state S_{i-1} emits a parton to produce the state S_i.
3. Reweight the events with $\prod_{i=3}^{n} \alpha_s(\rho_i) / \alpha_{\text{ME}}^{n-2}$.
4. For each state S_i (except S_n), generate an emission with ρ_i as starting scale and if this emission occurred at a scale larger than ρ_{i+1} reject the event. This is equivalent to reweighting with a factor $\prod_{i=2}^{n-1} \Delta_{S_i}(\rho_i, \rho_{i+1})$.
5. For the last step there are two cases.
   - If the event does not have the highest multiplicity $n < N$, generate an emission from the state S_n with ρ_n as starting scale. If the emission is above the merging scale Q_{MS}, reject the event. Otherwise accept the event and continue the cascade.
   - If the event has the highest possible multiplicity $n = N$, accept the event and start the cascade from the state S_n with the scale ρ_n.

For CKKW-L 1-jet merging to preserve unitarity, the first PS emission has to exactly match the 1-jet ME. This can be done (matching), but then there would be no need for 1-jet CKKW-L. For 2-jet merging, "the splitting kernels need to exactly reproduce the matrix element, phase space must be fully covered by the parton shower, and the no-emission probabilities need to be produced identically in both cases. Particularly the requirement that the phase space is completely covered is problematic, since parton showers commonly fill only phase space regions in which consecutive emissions are ordered in a decreasing evolution variable."[4]

[4] [L. Lönnblad, S. Prestel, 1211.4827]

UMEPS preserves the Born level cross section by integrating over $n+1$-jet phase space to induce resummation in n-jet cross section. The idea is to subtract in n what you add in $n-1$.

$$\sigma^{\text{inc}} = \int \mathrm{d}\phi_0 \left(\frac{\mathrm{d}\sigma_0^{\text{exc}}}{\mathrm{d}\phi_0} + \frac{\mathrm{d}\sigma_1^{\text{inc}}}{\mathrm{d}\phi_0} \right) \tag{1}$$

$$= \int \mathrm{d}\phi_0 \left(\frac{\mathrm{d}\sigma_0^{\text{inc}}}{\mathrm{d}\phi_0} - \frac{\mathrm{d}\sigma_{1\to 0}^{\text{inc}}}{\mathrm{d}\phi_0} + \frac{\mathrm{d}\sigma_1^{\text{inc}}}{\mathrm{d}\phi_0} \right) \tag{2}$$