

Simulating soft-QCD phenomena

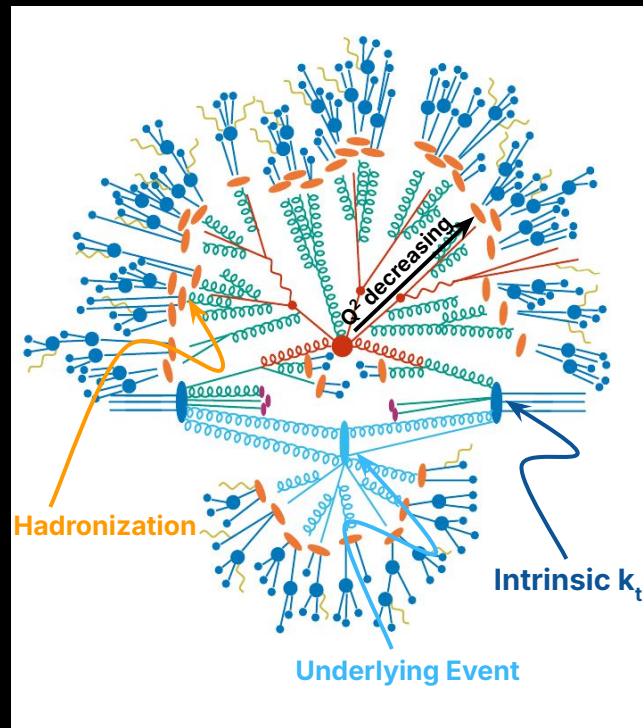
model calibration and uncertainty evaluations

Steffen Schumann

Institut für Theoretische Physik



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN IN PUBLICA COMMODO
SEIT 1737



Linking perturbative calculations to observations

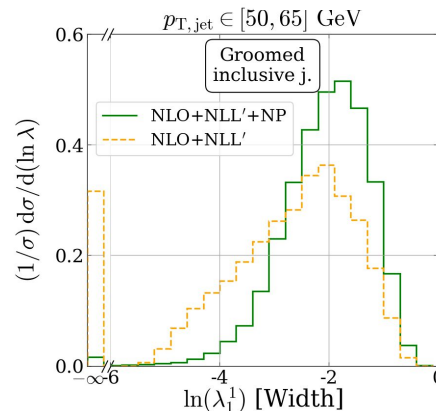
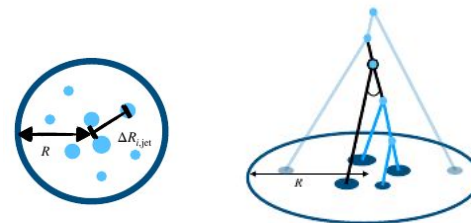
Soft-QCD phenomena omnipresent

- manifest for hadronic final states
 - From global event shapes to jet substructure observables
- unavoidable in hadronic collisions
 - Intrinsic transverse momentum
 - Underlying Event
- challenging physics questions
 - How do partons hadronize?
 - Fate of beam remnants?

Example: Resummed predictions for jet-angularities corrected with non-perturbative transfer matrices from MC

[Ghira et al. 2606.20798]

→ see talk by Andrea Mon @ 4:40pm



Pythia 8



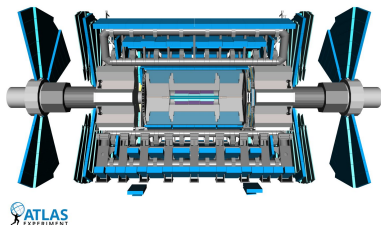
Herwig 7



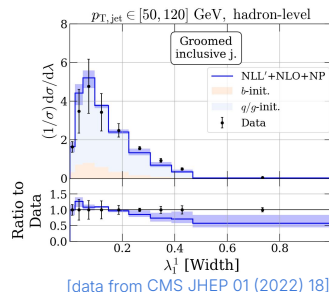
Sherpa 3

Computer models for complex physical processes

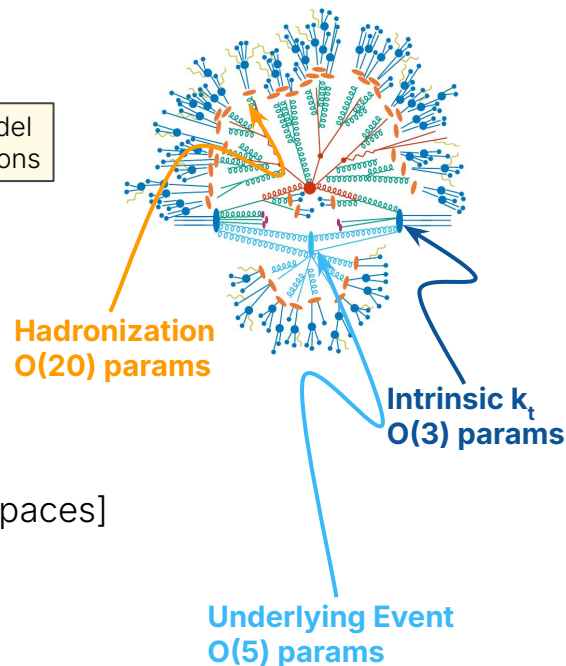
Contrast measured physical reality with (imperfect) model predictions



systematic errors
finite statistics



choice of model
model limitations



Spotlight on non-perturbative models

- specification of model(s) & input parameters prior to simulator run
 - dedicated calibration/tuning efforts [explore high-dimensional spaces]
 - needs accurate and sensitive measurement data
- quantification of systematic uncertainties
 - choice of concrete model/simulator, input-parameter variations
 - very costly to evaluate, aim for reweighting techniques

Modelling hadronization

Universality

- independence of very hard process [through parton shower]
- robust extrapolation to other processes, energies or colliders

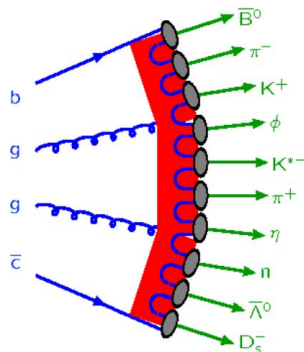
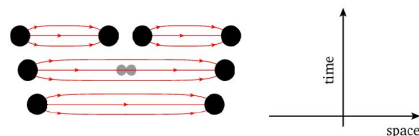
Modelling (un)known decays of (un)known hadrons

- hadron multiplicities, meson/baryon ratios
 - hadron-momentum distributions, event-shape variables, ...
 - re-analyses of archived LEP and HERA data very informative
 - [e.g. energy-correlators by Electron-Positron Alliance 2511.00149, (groomed) event shapes by H1 EPJC 84 (2024) 7/8]
- see talk by Yi Chen Tue @ 4:40pm

Lund-string model

PYTHIA

[Andersson et al. Phys Rept 97 (1983) 31]
[Bierlich et al. SciPost 97 2022 (2022) 8]



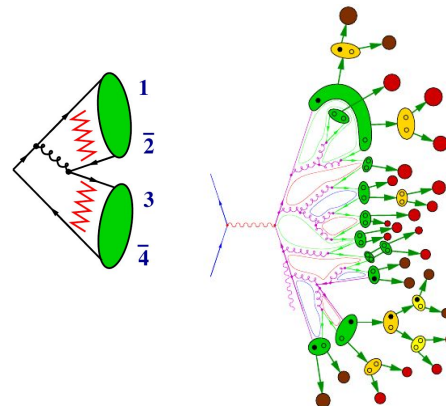
Cluster fragmentation

HERWIG

[Webber NPB 238 (1984) 492]
[Bellm et al. 2512.16645]

SHERPA

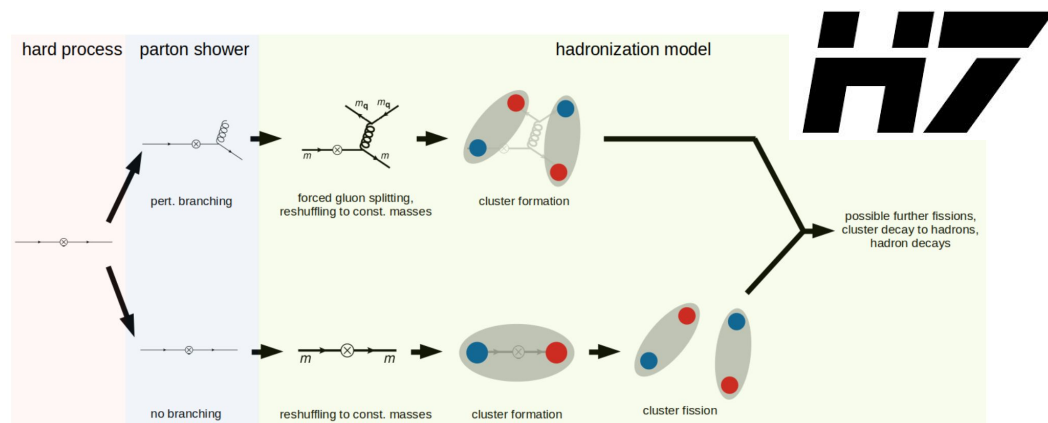
[Winter, Krauss EPJC 36 (2004) 381]
[Chahal, Krauss SciPost 13 (2022) 2]



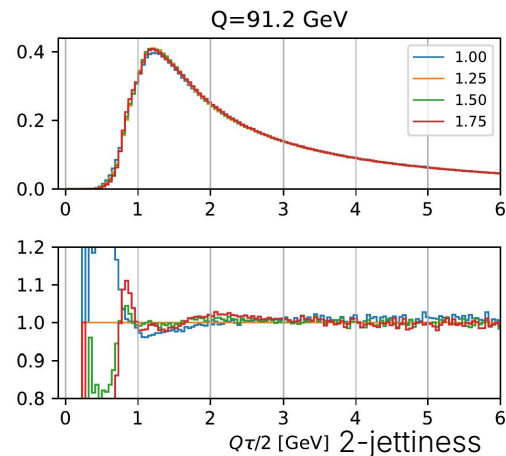
Modelling hadronization: initial conditions

Final state of perturbative event evolution constitutes input to non-perturbative phases

- sensitivity to infrared effects: higher-order corrections in hard process, details of shower algorithm
 - aim for perturbative modelling as accurate as possible
 - consider re-tuning when changing shower model or different hadronization for same shower
- allow for re-arrangement of large- N_c colour assignments, i.e. *Colour Reconnections*
- intriguing interplay of shower cut-off and hadronization-model parameters
 - interpret cut-off as IR factorization scale parameters can depend upon [\[Hoang et al. JHEP 07 \(2025\) 5\]](#)
 - towards factorization aware hadronization models



Reduced cut-off dependence
for new dynamical cluster model



Exploring high-dimensional spaces

Classic tuning strategy

“find (single) best-fit parameter point and possibly uncertainty”
based on *GoF measure*

guaranteed to find a solution

not necessarily unique, possibly unstable

History Matching approach [Iskauskas et al. JSS 109 (2024) 1]

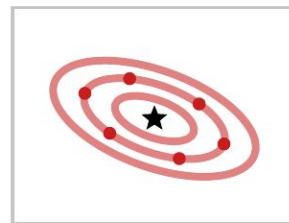
“identify parameter-space regions for which simulator output closely matches observations”

via complementarity to *implausible regions*

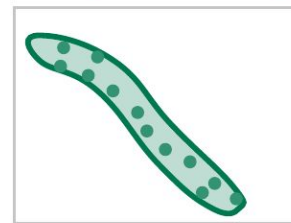
better captures multi-modality, correlations

requires samples from non-implausible region

Ellipsoid



Non-implausible region



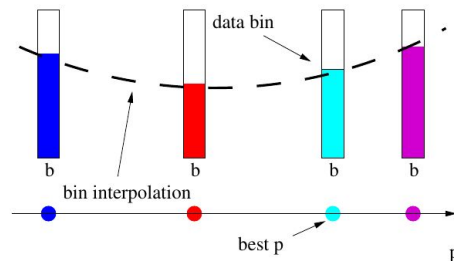
Commonalities: contrast input & observations

- select sensitive model outputs, i.e. observations
- use fast surrogate/emulator instead of simulator
- capture parameter dependence of model outputs

Professor/Apprentice approach

[Buckley et al. EPJC 65 (2010) 331]

[Krishnamoorthy et al. 2103.05748]

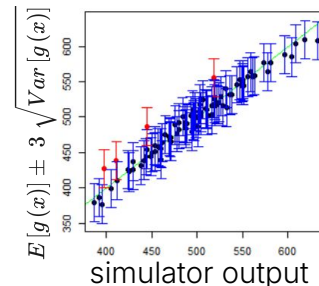
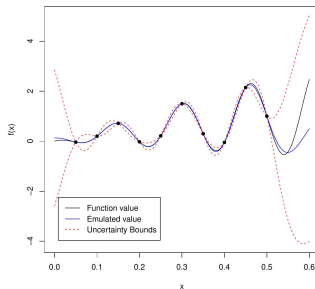
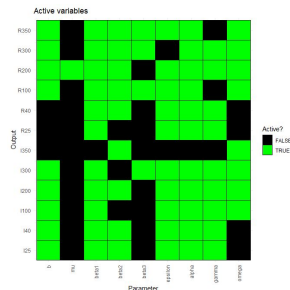


History Matching: Bayes Linear emulation

Construct fast emulator $g(x)$ for each considered observation (aka observable bin i)

$$g_i(x) = \sum_{j=1}^p \beta_{ij} h_{ij}(x) + u_i(x)$$

- known basis functions h_{ij} and stochastic process u_i (zero mean)
- use Bayes Linear statistics as approximation to full Bayes analysis [\[Goldstein, Wooff '07 Wiley & Sons\]](#)
 - emulator *expectation and variance as primitives* [no distributional statements needed]
 - possibly identify reduced set of *active variables* [\[Vernon et al. Bayesian Analysis 5 \(2010\) 619\]](#)
 - update beliefs about $E[g_i(x)]$ and $Var[g_i(x)]$, diagnostic analysis using validation runs



History Matching: implausibility measure

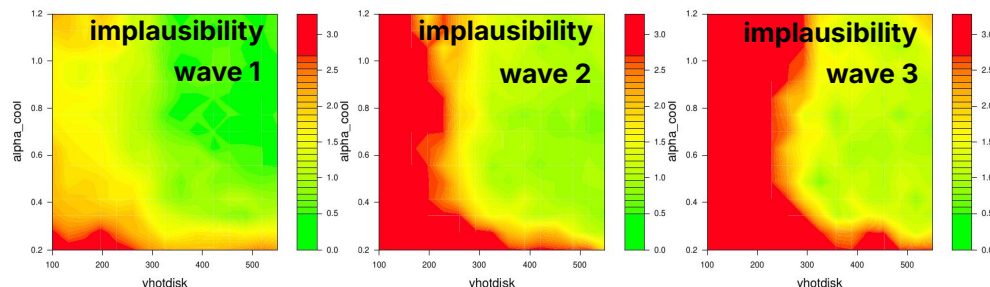
Exclude implausible parameter regions based on Bayes Linear emulators

- consider actual observation (data point) $\mathcal{Z}_i$ with uncertainty $Var[e_i]$
- Implausibility measure** for parameter point based on emulator expectation & variance

$$I_i^2(x) = \frac{(E[g_i(x)] - z_i)^2}{Var[g_i(x)] + Var[e_i] + Var[\epsilon_i(x)]}$$

- $Var[\epsilon_i(x)]$ accounts for additional modelling uncertainty [expert elicitation]
- discard input-parameter point if $I_M(x) = \max_{i=1, \dots, N} I_i(x) > 3$

- consecutive waves of History Matching
 - sample from non-implausible regions
 - train & validate new emulators
 - compress parameter-space volume



HM: hadronization-model calibration

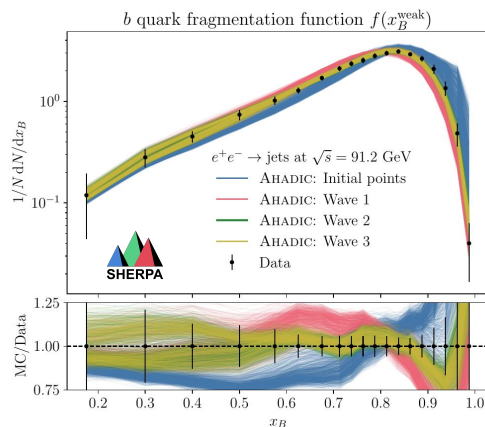
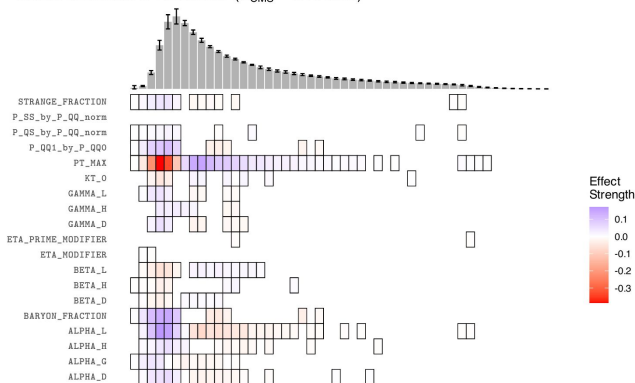
Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA

[Iskauskas et al. JHEP 08 (2026) 066]

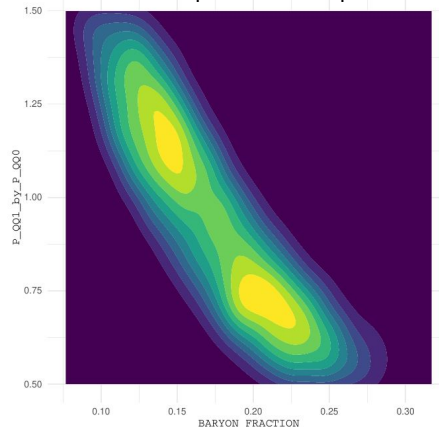
- identical perturbative setup $e^+e^- \rightarrow qq+\{0,1\}j$ @ NLO + SHERPA dipole shower
 - facilitates assessment of hadronization-model dependence
- consider 19 parameters for AHADIC, 23 for PYTHIA Lund-string model
- constrained using large set of LEP1 data, e.g. event shapes, hadron multiplicities, ...
 - total of 432 bins, determine reduced sets via Principal Variable Analysis

[Cumming, Wooff Computational Statistics & Data Analysis 52 (2007) 550]

Active Variables: C-Parameter ($E_{\text{CMS}} = 91.2\text{GeV}$)



AHADIC: 2D projection final non-improbable space



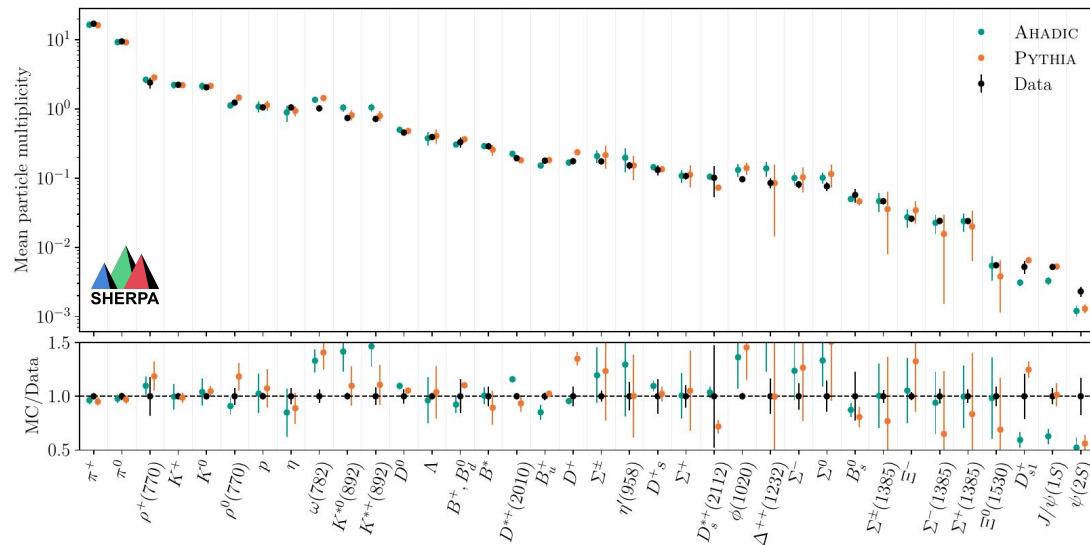
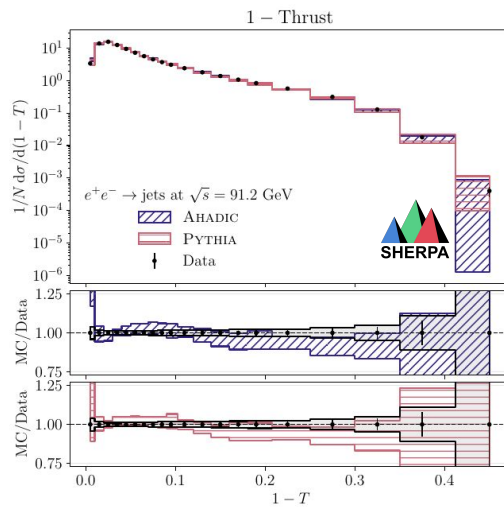
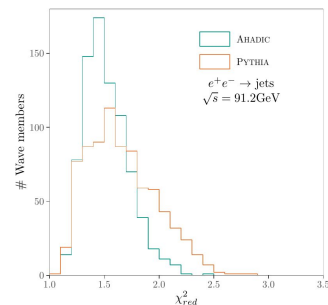
HM: hadronization-model calibration

Cluster- and Lund-string (via PYTHIA 8) fragmentation for SHERPA

[Iskauskas et al. JHEP 08 (2026) 066]

- samples from final non-implausible space provide uncertainty estimate
 - here considered space-filling O(800) parameter sets
 - reflect parameter uncertainties, similar quality of data description

χ_{red}^2 full set of 432 outputs



Parameter Variations: Reweighting Techniques

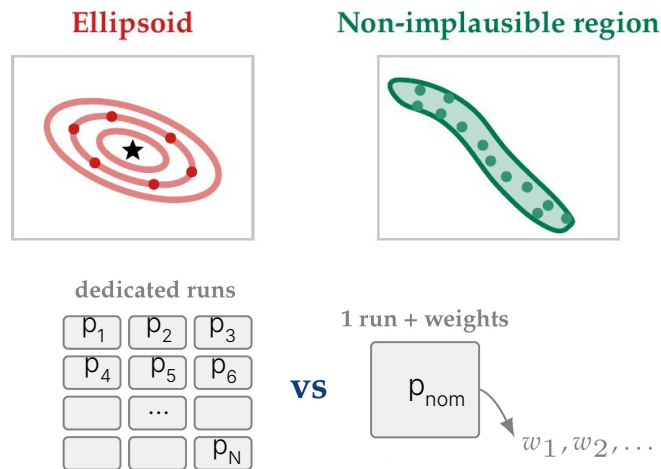
- independent of calibration method, **Professor/Apprentice** or **History Matching**, fitting procedure and uncertainty evaluation require *many generator runs*
- Reweighting Techniques** established standard for the evaluation of perturbative uncertainties: μ_R , μ_F , α_s , PDFs
→ represented by event-weight variations
[Bellm et al. PRD 94 (2016) 3] [Mrenna, Skands PRD 94 (2016) 7]
[Bothmann et al. EPJC 76 (2016) 11]
- reweighting non-perturbative model-parameter variations
→ event-wise book-keeping of decisions

In-situ approach (on-the-fly variation-weight evaluation)

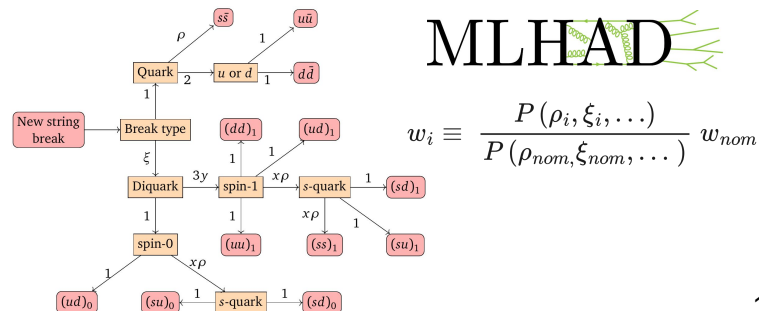
- MLHad collaboration for string-fragmentation function
[Bierlich et al. SciPost 16 (2024) 5]
- Underlying Event & Colour Reconnections in SHERPA
[Pabst et al. JHEP 09 (2026) 87]

Post-hoc approach (a-posteriori weight re-calculation)

- MLHad for flavour-selection in string model
[Assi et al. SciPost 19 (2025) 4]



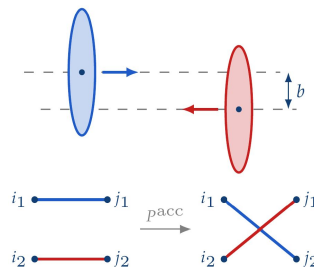
Example: Flavour selection string-breaking (simplified)



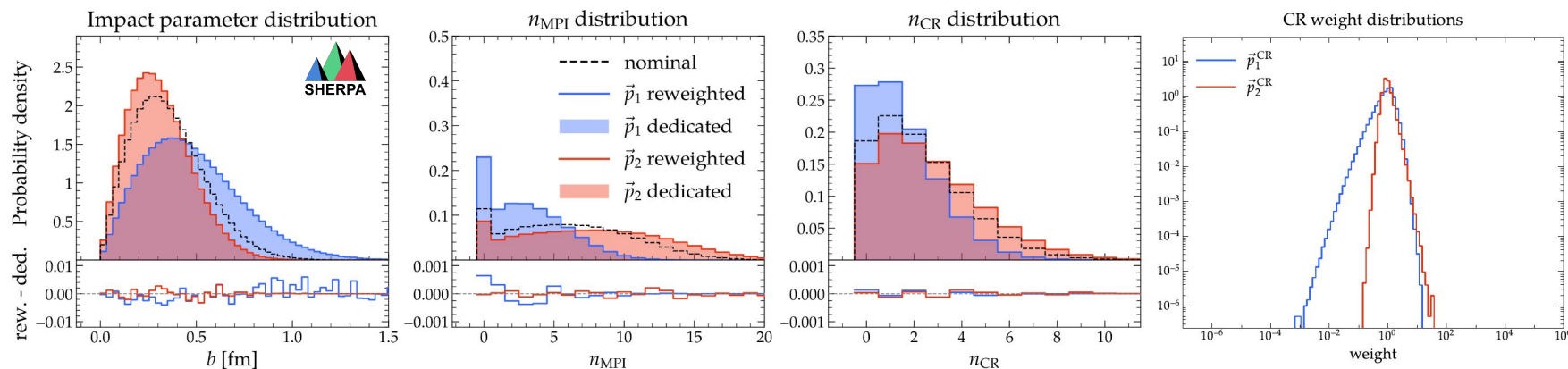
Parameter Variations: Reweighting Techniques

On-the-fly reweighting of UE and Colour Reconnections in SHERPA [Pabst et al. JHEP 09 (2026) 87]

- modelling Multi-Parton-Interactions: Gaussian hadronic-matter profiles [Sjöstrand, Ziji PRD 36 (1987) 2019]
- swaps of colour partners, based on momentum-space distance
→ enable in-situ variations of all relevant model parameters
→ check closure of reweighted & dedicated predictions

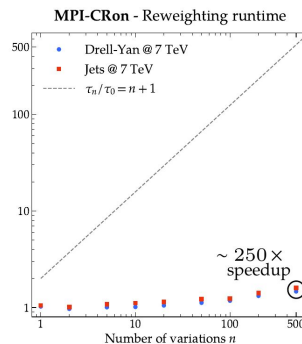
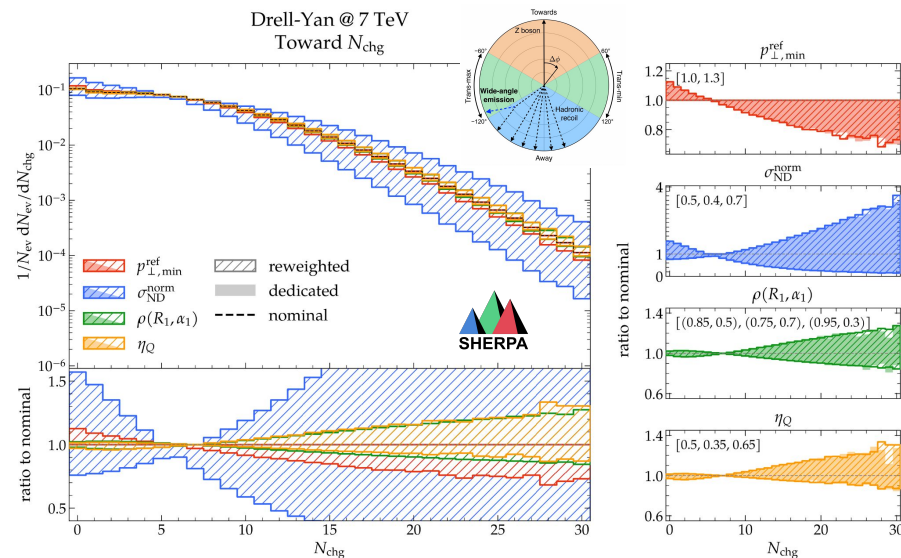


Drell-Yan @ 7 TeV



Parameter Variations: Reweighting Techniques

On-the-fly reweighting of UE and Colour Reconnections in SHERPA [Pabst et al. JHEP 09 (2026) 87]



Application: MPI+CR tuning using Apprentice

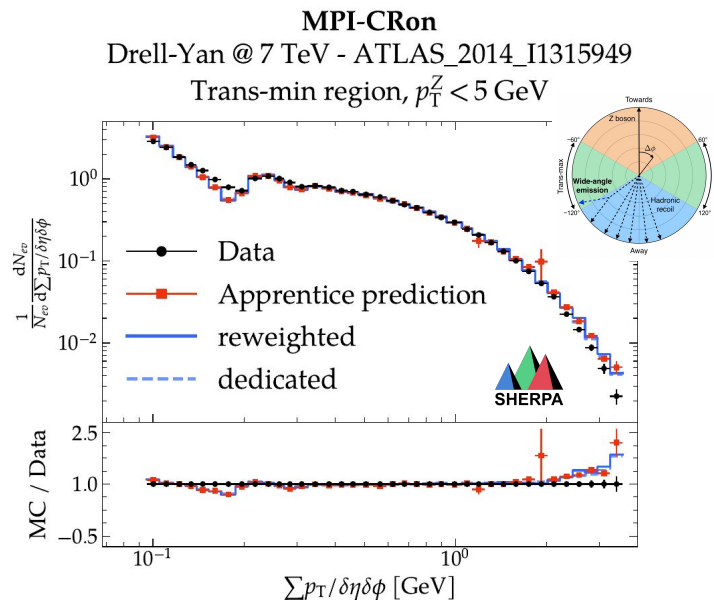
- using 7 TeV Drell-Yan and pure Jets data
- grid-point generation via reweighting

500 param sets	dedicated runs	reweighting
generator CPU	~500x	~2x
sample storage	500	1
post-processing	500	1

Parameter Variations: Reweighting Techniques

On-the-fly reweighting of UE and Colour Reconnections in SHERPA [Pabst et al. JHEP 09 (2026) 87]

- perform separate MPI tunes for CR on/off, dubbed MPI-CR_{on}/MPI-CR_{off}
→ considered data favours CR enabled



Process/Analysis	χ^2/ndf		ndf
	MPI-CR _{off}	MPI-CR _{on}	
Drell-Yan (cum.)	34.0	16.1	2573
ATLAS_2014_I1315949	32.4	14.7	2123
CMS_2012_I1107658	41.7	22.6	450
Jets (cum.)	2.5	2.0	1690
ATLAS_2014_I1298811	2.7	2.1	1035
ATLAS_2011_I929691	1.6	1.4	409
CMS_2011_I916908	3.6	2.1	163
CMS_2011_I886332	3.1	4.0	78
CMS_2011_I930319	1.1	1.2	5

- MPI-CR_{on} tune new default since v3.0.2
- MPI+CR reweighting to be released with v3.1
- Cluster-model reweighting work in progress

Simulating soft-QCD phenomena

Analysis of non-perturbative uncertainties

- Careful calibration of model parameters (aka tuning)
- Quantification of model-inherent uncertainties
- Estimate dependence on choice of model
→ thanks to modularity of modern event generators

First application of History Matching for hadronization-model calibration

- compress parameter space based on Bayes Linear emulators
- samples from non-improbable space provide uncertainty estimate

Reweighting Techniques for non-perturbative models

- resource-efficient evaluation of model-parameter variations
- novel applications from tuning, uncertainty quantification to sensitivity studies

Many more (specific) model developments, e.g.

- rework cluster model in Herwig [[Gieseke et al. EPJC 85 \(2025\) 7](#)]
- quark-spin effects in Lund-string model [[Kerbizi, Lonnblad 2606.27162](#)]
- new, data driven, ML/AI approaches to hadronization [[MLHad and HadML collaborations](#)]
→ see talk by Tilman Fri @ 10am

