

Modelling of LHC Collisions in PYTHIA

— Physics & Uncertainties

Peter Z Skands

& Monash University University of Oxford

 \mathbf{A} big than \mathbf{A} lovely feedback following the last Oxford, April 2024

FSR

- Hard Process
- O Hard Interaction
- **•** Resonance Decays
- **MECs, Matching & Merging**

FSR

Recent Studies

Focus on SM precision environments \leftrightarrow BSM backgrounds

-
- 1. NLO Matching Systematics with POWHEG-Box (examples: VBF, tt) 2. From NLO to **NNLO** (examples: tt̄, V, H, VH, VV, ...)
- 3. The computational bottleneck in ME merging *(example: V+jets)*
- 4. New Discoveries in Hadronization *(examples: HF baryons, JES)*

NB: want to address/explain state of the art & systematics in real contexts \rightarrow a bit theory heavy

1. NLO + Shower with POWHEG

 $\frac{1}{2}$ Pseudorapidity of the emitted parton

It does so in a shower-like manner: sweeping over phase space, from high to low p_T

Matrix-Element Corrections (MECs) [Bengtsson & Sjöstrand 1987 + …]

POWHEG generates the first (hardest) emission with $\vert M_{X+1} \vert$ 2

+ NLO Born Normalization [Nason 2004; Fixione, Nason, Oleari 2007]

Shower then takes over and generates all further emissions

Using soft/collinear approximations

(Just focusing on the real-radiation part)

Nason 2004; Fixione, Nason, Oleari 2007

Generic emission phase space

Phase Space already
Covered by Powheg Powheg Emission
Generated with $|M_{X+1}^{\text{lo}}|^2$ -Phase Space $\frac{1}{\frac{1}{x}}$ // $\frac{1}{\frac{1}{x}}$ /// $\frac{1}{\frac{1}{x}}$ //shower

Powheg Box — A Subtlety

Δ

171

Can be important for complex / multi-scale processes.
 VBF: Höche et al., SciPost Phys. [12 \(2022\) 1](https://arxiv.org/abs/2106.10987)

[Alioli et al, 2010]

- Exploits having its own definition of " p_T "
- \neq shower's definition of p $_{\sf T}$
- Breaks clean matching

E.g., Nason, Oleari [a](https://arxiv.org/abs/2106.10987)[rXiv:1303.3922](https://arxiv.org/abs/1303.3922)

Industry Standard: "Powheg Box"

Solution: Vetoed Showers

- (+ truncated showers)
- Works very well for simple cases

Induces an uncertainty/ambiguity

Purely associated with the matching scheme (not physical)

A More Complex Process

Vector boson fusion, *qq* → *q*′*q*′*H*

Again, POWHEG-Box generates the first emission, which it judges to be the "hardest" according to its own p_T definition

 p_\perp with respect to the beam p_\perp with respect to the final-state q' partons p_\perp with respect to either of the (q_*q') dipoles q^*q p_\perp with respect to the H ? (+ PYTHIA defines a problematic $(q'q')$ dipole) + Interpolations/combinations of the above … crossed

Note: similar concerns for any process with coloured partons in the final state at Born level $t\bar{t}$ (& $t \to bW$), $V/H + \text{iet(s)}$, director. $t\bar{t}$ (& $t \rightarrow bW$), $V/H + \text{jet}(s)$, dijets, trijets, ...

Multiple emitters \rightarrow several overlapping phase spaces

And many possible p_T definitions:

POWHEG-Box Matching Systematics china Systematics

h

$\overline{8.5}$

Pseudorapidity Difference of the Fourth and Tagging Jets

1 2 $\overline{25}$ Ratio

2. From NLO to NNLO

MiNNLO*PS* builds on (extends) POWHEG NLO for X + jet

- Allow the first jet to approach $p_\perp \rightarrow 0 \sim X + 0$
- Tame divergence with analytic (NNLL) Sudakov \cdot
	- (introduces additional hardness scale = resummation scale)
- **Normalize** inclusive $d\sigma_X$ to NNLO
	- $(\mathcal{O}(\alpha_s^3))$ ambiguity on how to "spread" the additional contributions in phase space.)

~ Best you can do with current off-the-shelf parton showers off-the-shelf parton showers

Is approximate; introduces some ambiguities: p_\perp^Shower vs p_\perp^Powheg vs $\mathcal{Q}_{NNLL}^\text{resummation}$ & differential NNLO spreading (+ possible efficiency bottleneck: $p_\perp \rightarrow 0$ singularity \times Sudakov veto)

What if we could lift that restriction?

[Hamilton et al. 1212.4504, Monni et al. 1908.06987]

MiNNLOPS inherits some issues from POWHEG-Box applicant [88], constituting the **first-ever and so-far only approach of its kind**. Marinelli, Riccardo Nagar, and Davide Napoletano. Double Higgs production at NNLO interfaced to parton showers in IGGU*RG* TROM PUWHFG-

Large dependence on **pThard** scale

Big variations in predictions for further jets

 \implies Fairly big variations for Born-level (0-jet) observable.

Calculation "anchored" in NLO for X+jet

approximate NNLO matching scheme. Uncertainties of other sources

Recommendations to Users of these Calculations

MiNNLO*PS* is an *approximate* matching scheme ~ Best you can do with current off-the-shelf parton showers! But: does not "match" shower to NNLO point by point in phase space (Impossible to do so with LL showers.) Does not (always) do vetoed showers (This can in principle be done.) Depends on several auxiliary scales (Intrinsic to scheme. Physical observables should not depend on them → *vary!*)

Do comprehensive variations to estimate scheme uncertainties Subsequent shower not fully *guaranteed* to preserve accuracy (Also applies to POWHEG + showers)

-
-

Towards True* NNLO Matching

*In the sense of the fixed-order and shower calculations matching each other point by point in each phase space

Idea: Use (nested) Shower Markov Chain as NNLO Phase-Space Generator

- Harnesses the power of showers as efficient phase-space generators for QCD
	- Pre-weighted with the (leading) QCD singular structures = soft/collinear poles

Different from conventional Fixed-Order phase-space generation (eg VEGAS)

Continue shower afterwards

- No auxiliary / unphysical scales
	- ⇒ expect small matching systematics

Towards True* NNLO Matching

*In the sense of the fixed-order and shower calculations matching each other point by point in each phase space

Need:

- **D** Born-Local NNLO $(\mathcal{O}(\alpha_s^2))$ K-factors: $k_{NNLO}(\Phi_2)$
-
-

⁰ *‡*(0)

⁰ *‡*(0)

¹ *‡*(0)

² *‡*(0)

3 ...

Recent Studies

Focus on SM precision environments \leftrightarrow BSM backgrounds

- 1. NLO Matching Systematics with POWHEG-Box (examples: VBF, tt)
- 2. From NLO to **NNLO** (examples: tt̄, V, H, VH, VV, ...)
- 3. The computational bottleneck in ME merging *(example: V+jets)*
- 4. New Discoveries in Hadronization *(examples: HF baryons, JES)*

NB: want to address/explain state of the art & systematics in real contexts \rightarrow a bit theory heavy

Something about CPU resources c en Condensed rem Condensed remarks from talk by T. Moskalets (ATLAS) at CERN Workshop Nov 2023

The Computational Bottleneck in ME Merging the conservative (blue) and aggressive (red) R&D scenarios

Tetiana Moskalets | Event generators' and N(n)LO codes' acceleration | 13-14 Nov 2023 **‣ Largest fraction of EvGen CPU time is taken by generation of multi-leg MC predictions**

namely, multijet merged Sherpa V+jets

Year

Each phase-space point receives contributions from many possible branching "histories" (aka "clusterings") \overline{c}

Need to take all contributing "shower histories" into account.

In conventional parton showers (Pythia, Herwig, Sherpa, ...)

Matrix-Element Merging – The Complexity Bottleneck

For CKKW-L style merging: (incl UMEPS, NL3, UNLOPS, ...) (*t*0*,t*)

of histories grows ~ # of Feynman Diagrams, **faster than factorial** number of histories **scales factorially** with number of legs

Bottleneck for merging at high multiplicities (+ high code complexity) sections of morging acting minateprotect (!) ingli code com

-
-

Sector Showers (without maths)

VINCIA's shower is unique in being a "Sector Shower" Partition N-gluon Phase Space into N "sectors" (using step functions). Each sector \leftrightarrow one specific gluon being the "softest" in the event Inside each sector, only one kernel contributes (the most singular one)! Sector Kernel = the eikonal for the soft gluon and its collinear DGLAP limits for $z > 0.5$. → Unique properties: shower operator is bijective and is a true Markov chain

- PS & Villarejo [JHEP 11 \(2011\) 150](https://arxiv.org/abs/1109.3608) Brooks, Preuss, PS [JHEP 07 \(2020\) 032](https://arxiv.org/abs/2003.00702)
-
-
-
-
-

The crucial aspect:

Only a single history contributes to each phase-space point !

\Longrightarrow Factorial growth of number of histories reduced to constant! (And the number of sectors only grows linearly with the number of gluons) $(g \to q\bar{q} \to$ leftover factorial in number of same-flavour quarks; not a big problem)

Sectorized CKKW-L Merging publicly available from Pythia 8.306

Brooks & Preuss, "Efficient multi-jet merging with the VINCIA sector shower", [arXiv:2008.09468](https://arxiv.org/abs/2008.09468)

Demonstrated constant scaling with multiplicity. Extensions now pursued:

Optimisations of baseline algorithm Figure 14: PYTHIA and VINCIA CPU time scaling in history construction (*left*) and parton-level event generation (*right*) for \mathbf{A} as a gauge of the scaling behaviour of the memory usage in both merging in both merging in both memory usag

> Sectorized iterated tree-level ME corrections (demonstrated in PS & Villarejo arXiv[:1109.3608\)](https://arxiv.org/abs/1109.3608) statistically pornon didependent runs and from 7 jets on, whi<u>le provisions,</u>

Sectorized multi-leg merging at NLO (active research grants, with C. Preuss, Wuppertal) strategies to deal with competing sectors, cf. e.g. [68, 70], which can improve the performance relative to de
The performance relative to deal with can improve the performance relative to deal with can improve the perform the VINCIA curve remains almost flat, with only a small peak around 3 additional jets. The latter can be latter
The latter can be latter c

Recent Studies

Focus on SM precision environments \leftrightarrow BSM backgrounds

- 1. NLO Matching Systematics with POWHEG-Box (examples: VBF, tt)
- 2. From NLO to **NNLO** (examples: tt̄, V, H, VH, VV, ...)
- 3. The computational bottleneck in ME merging *(example: V+jets)*

4. New Discoveries in Hadronization *(examples: HF baryons, JES)*

NB: want to address/explain state of the art & systematics in real contexts \rightarrow a bit theory heavy

LHC experiments report very large (factor-10) enhancements in heavyflavour baryon-to-meson ratios at low p_T!

New Discoveries in Hadronization

Not predicted by default Pythia (Monash)

Very exciting!

26

 $q\bar{q}$ strings (with gluon kinks) E.g., $Z \rightarrow q\bar{q}$ + shower $H \rightarrow b \bar{b} + {\rm shower}$

What are **"String Junctions"**?

27

Closed Strings

SU(3) String Junction

Gluon rings E.g., $H \rightarrow gg +$ shower $\Upsilon \rightarrow ggg$ + shower Open strings with $N_C = 3$ endpoints E.g., Baryon-Number violating ${\mathfrak{n}}$ eutralino decay ${\tilde{\chi}}^0 \to q \overline{q} q + {\mathsf{shower}}$

QCD Colour Reconnections How do Colour Reconnections Create String Junctions?

Stochastically restores colour-space ambiguities according to **SU(3) algebra**

 \triangleright Allows for reconnections to minimise string lengths

What about the **red-green-blue** colour singlet state?

Junctions!

[Christiansen & PS [JHEP 08 \(2015\) 003\]](https://arxiv.org/abs/1505.01681)

What do String Junctions do?

Assume Junction Strings have same properties as ordinary ones (u:d:s, Schwinger p_T , etc)

The Junction Baryon is the most "subleading" hadron in all three "jets".

Generic prediction: low p_T

A Smoking Gun for String Junctions: Baryon enhancements at low p_T

[Sjöstrand & PS, [NPB 659 \(2003\) 243\]](https://arxiv.org/abs/2309.01557) [+ J. Altmann & PS, in progress]

What a **strange** world we live in, said Alice

We also know ratios of strange hadrons to pions strongly increase with event activity − πVi
I ່ to
∙ article
and the contribution
of the contribution of the con
of the contribution of the contribution of the contribution of the contribution of π + $\frac{1}{\sqrt{2}}$ it
I π + πRatio of yields to (

TOPOLOGICAL PHOTONICS Optical Weyl points and Fermi arcs

31

In Progress: Strangeness Enhancement from Close-Packing Idea: each string exists in an effective background produced by the others ness Enhancement Strangene

-
-

Particle Composition: Impact on Jet Energy Scale

ATL-PHYS-PUB-2022-021

ATLAS PUB Note

29th April 2022

[Dependence of the Jet Energy Scale on the Particle](https://cds.cern.ch/record/2808016/files/ATL-PHYS-PUB-2022-021.pdf) of the Particle on the Particle on the Particle Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation

Content of Hadronic Jets in the ATLAS Detector Simulation in Monte Carlo simulations of the particle types and spectra within jets is $\frac{1}{5}$ to the true jet energy kaons and baryons in the jet. Model differences observed for jets initiated by *quarks* or *gluons* produced in the hard scattering process are $\frac{1}{\sqrt{1}}$ dominated by the differences in these hadron energy fractions indicating that **measurements of the hadron content of jets and improved tuning** of hadronization models can result in an improvement in the precision $\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1$ of the knowledge of the ATLAS jet energy scale. The dependence of the ATLAS jet energy measurement on the modelling investigated. **It is found that the hadronic jet response, i.e. the ratio of** the reconstructed jet energy to the true jet energy, varies by $\sim 1-2\%$ **depending on the hadronisation model used in the simulation. This effect is mainly due to differences in the average energy carried by**

- Variation largest for gluon jets For $E_T = [30, 100, 200]$ GeV Max JES variation = $[3\%, 2\%, 1.2\%]$
- Fraction of jet E_T carried by baryons (and kaons) varies significantly
	- Reweighting to force similar baryon and kaon fractions
	- Max variation → [1.2%, 0.8%, 0.5%]
	- Significant potential for improved Jet Energy Scale uncertainties!
- Motivates Careful Models & Careful **Constraints**
	- Interplay with advanced UE models
	- In-situ constraints from LHC data
	- Revisit comparisons to LEP data

Summary & Outlook

State of the art for perturbation theory: NNLO (**→** N3LO)

Showers + hadronization mandatory for collider studies (+ resummation extends range)

Now: can use off-the-shelf showers with MiNNLO*PS*

Based on POWHEG-Box + Analytical Resummation + NNLO normalisation Approximate method; depends on several auxiliary unphysical scales → can exhibit large variations

Work in progress: VinciaNNLO

New research grant with LHCb (Warwick) focusing on strings with b-quark endpoints And QED corrections in B decays

-
-

Based on nested shower-like phase-space generation with second-order MECs True NNLO matching \rightarrow Expect small matching systematics So far only worked out for colour-singlet decays. (Also developing extensions towards NLL, NNLL showers …)

Beautiful Strings

New discoveries at LHC on particle composition, esp. **baryons and strangeness**

Extra Slides

Parton Showers: Theory

Mathematically, **gauge amplitudes factorize** in **singular limits**

Partons ab $|\mathcal{M}_{F+1}(\ldots, a, b, \ldots)|$
 \rightarrow **collinear**: $2 \frac{a||b}{ }$ $\overset{a||b}{\rightarrow} g_s^2 \mathcal{C} \frac{P(z)}{2(n-r)}$

 $P(z) =$ **DGLAP** splitting kernels", with $z = E_a / (E_a + E_b)$

$$
\frac{P(z)}{2(p_a \cdot p_b)}|{\cal M}_F(\ldots,a+b,\ldots)|^2
$$

$$
\underset{\theta}{\text{Gluon j}} \quad |\mathcal{M}_{F+1}(\ldots,i,j,k\ldots)|^2 \stackrel{j_g \to 0}{\to} g_s^2 \mathcal{C} \frac{(p_i \cdot p_k)}{(p_i \cdot p_j)(p_j \cdot p_k)} |\mathcal{M}_F(\ldots,i,k,\ldots)|^2
$$

These are the **building blocks of parton showers** (DGLAP, dipole, antenna, ...) (+ running coupling, unitarity, and explicit energy-momentum conservation.)

Coherence → Parton j really emitted by (i,k) "dipole" or **"antenna" (eikonal factors)**

see e.g PS, *Introduction to QCD*, TASI 2012, [arXiv:1207.2389](http://arxiv.org/abs/arXiv:1207.2389)

Most bremsstrahlung is driven by divergent propagators → simple structure

Confinement in PYTHIA: *The Lund String Model*

Simplified (leading-Nc) "colour flow" \rightarrow determine between which partons to set up confining potentials

Map from Partons to Strings:

Quarks ➡ string endpoints; gluons ➡ transverse "kinks"

System then evolves as a string world sheet

+ **String breaks** via spontaneous $q\bar{q}$ pair creation ("Schwinger mechanism") \rightarrow hadrons

The String Fragmentation Function

Consider a string break ω , producing a meson M, and a leftover string piece Siuci a sumy D
A niere

The meson M takes a fraction z of the quark momentum,

Lorentz invariance \Longrightarrow string breaks can be considered in *any order*. Imposes "left-right symmetry" on the FF

 \Longrightarrow FF constrained to a form with two free parameters, a & b : constrained by fits to measured hadron spectra

Supresses high-z hadrons

Supresses low-z hadrons

Observation: All string breaks are causally disconnected

1 *z* $(1 - z)^a$ $\exp\left(-\frac{b(m_h^2 + p_{\perp h}^2)}{z}\right)$ *z*)

Automated Hadronization Uncertainties

Problem:

Given a colour-singlet system that (randomly) broke up into a specific set of hadrons:

- What is the **relative probability** that same system would have resulted, if the fragmentation parameters had been different? Would this particular final state become **more likely** ($w' > 1$)? Or **less likely** ($w' < 1$) Crucially: maintaining unitarity \Longrightarrow inclusive cross section remains unchanged!
- August 2023: Bierlich, Ilten, Menzo, Mrenna, Szewc, Wilkinson, Youssef, Zupan [*Reweighting MC Predictions & Automated Fragmentation Variations in Pythia 8*, [2308.13459\]](https://arxiv.org/abs/2308.13459)
	- Method is general; demonstrated on variations of the 7 main parameters governing longitudinal and transverse fragmentation functions in PYTHIA 8
- <https://gitlab.com/uchep/mlhad-weights-validation>

39

modification *zrQbm*²

 \mathbf{C}

A Brief History of MPI in PYTHIA

