May 2026

Implementation of Novel NLO Matching Schemes for $pp \to Z^0 Z^0$

Yuxiao Che

Theoretical Particle Physics
Division of Particle and Nuclear Physics
Department of Physics
Lund University

Master thesis (60 hp) supervised by Rikkert Frederix

Abstract

The task of matching fixed-order event generation with the parton shower is physically non-trivial and closely related to an efficiency bottleneck of event generators. In this thesis, we examine and implement two novel matching schemes in KrkNLO and MAcNLOPS for the LHC process $pp \to Z^0 Z^0$. As both schemes generate positive-definite event samples, they can be interesting alternatives to MC@NLO. The implementation of KrkNLO will be modified compared to original formulation, allowing $\overline{\text{MS}}$ PDFs to be used. Whereas for MAcNLOPS, this will be its first implementation. Both implementations rely on MADGRAPH5_AMC@NLO (MG5_AMC) and PYTHIA8 for hard event generation and parton shower respectively, using also elements of MG5_AMC for the matching procedure. The results are obtained for various distributions sensitive to matching schemes or valuable for practical purposes. In both cases, the results agree well with MC@NLO up to the higher order effects. For KrkNLO, it was additionally observed to be insensitive to factorisation scheme choice.

Popular Science Summary

Throughout the history of physics, the dominant philosophy had generally been inclined towards reductionism. This belief in the possibility of understanding complex phenomena in terms of basic constituents motivated the ultimate goal of finding truly fundamental laws of nature. An incredible advancement towards this goal took place in the past century, where physicists managed to identify the most basic constituents of matter, in the form of fundamental particles, and proposed a theory describing their interactions. This theoretical framework with the ambitious name of the *Standard Model* is the most successful theory of known elementary particles and their non-gravitational interactions.

However, despite being powerful and conceptually elegant, using the Standard Model entails extremely complicated calculations, posing a huge challenge for connecting theory with experiment. In fact, the data obtained from detectors at collider experiments are cannot be directly interpreted without sophisticated reconstruction and simulation tools. To translate them to physically meaningful results, computer simulations must be used. These simulations start from theory and give the expected results at an experiment, hence can be used backwards to comprehend the actual data. They are known as *event generators* and are now indispensable tools in particle physics.

As event generators are so important, their efficiency is a huge concern, especially considering the pending upgrades to collider experiments. Take the High-Luminosity Large Hadron Collider as an example. When its construction is complete in 2030, the data it is expected to collect per year is 1.5 times the total data collected in the entire three years of run 2. This is a great challenge on the computation front, both in terms of CPU usage and storage capacity.

Unfortunately, bottlenecks on the efficiency of event generators exist due to physical constraints. One of the more prominent bottlenecks is associated with weights of the events generated. If the weights are non-uniform, we will get greater imprecision. Since the requirement is typically put on the precision of results, a greater sample size is needed, leading to more CPU time and storage.

Up to the common target of producing the accurate result, these weights can be manipulated by the usage of different *matching schemes*, with the standard scheme giving an event sample uniform in magnitude but with a fraction being negatively weighted. Recently, two schemes taking the names `KrkNLO` and `MAcNLOPS` were proposed. They are closely related to the standard scheme, but have the advantage of fully removing the negative weights with negligible downside. This is a solid step in the right direction of having event samples with uniform and positive weights.

My work is centred on a new take on the implementation of the two schemes, providing fresh insights into their performance and future developments. The results obtained are very promising, generally presenting them as plausible alternatives to the standard scheme. If further tests continue to deliver similar results, a discussion about replacing the standard scheme with one of them might take place in the future. Considering the fact that `MAcNLOPS` is already capable of producing positive and uniform event samples, its wide implementation may represent a final resolution to this computational bottleneck.

Contents

1 Introduction **2**

2 Background **3**
2.1 Fixed-Order NLO . . . 3
2.2 Parton Showers . . . 4
2.3 Phase Space Mappings . . . 6

3 NLO Matching **8**
3.1 No Matching . . . 8
3.2 MC@NLO . . . 9
3.3 KrkNLO . . . 12
3.4 MAcNLOPS . . . 14

4 Implementation **16**
4.1 pKrkNLO with MG5_aMC . . . 16
4.2 MAcNLOPS with MG5_aMC . . . 17
4.3 Common Technicalities . . . 17

5 Results **18**
5.1 pKrkNLO . . . 18
5.2 MAcNLOPS . . . 23

6 Conclusion **28**

A Transformation to Krk PDFs **30**

B Event Samples with Different Weight Structures **34**

List of Acronyms

1. LO - Leading Order
2. NLO - Next-to-Leading Order
3. NLL - Next-to-Leading Logarithmic
4. LHC - Large Hadron Collider
5. PS - Parton Shower
6. ISR - Initial-State Radiation
7. FSR - Final-State Radiation
8. PDF - Parton Distribution Function
9. MG5_aMC - MadGraph5_aMC@NLO

List of Figures

1 Illustrations of (a) FSR and (b) ISR cascades 5
2 Example of a histogram for transverse momentum of the leading jet, with contributions from $\mathbb{S}$- and $\mathbb{H}$-events shown separately. Negative bins are reflected and labelled using dotted lines. 11
4 Histograms for the leading jet transverse momentum distribution after first emission obtained with `pKrkNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference. . . 19
5 Histograms for distribution of difference in azimuthal angle of Z^0-pair after (a) the first emission and (b) full shower for `pKrkNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference. 20
6 Histograms for the leading jet rapidity distribution after first emission obtained with `pKrkNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference. 21
7 Histograms for (a,b) invariant mass and (c,d) rapidity separation of the Z^0-pair after first emission (left) and full shower (right) obtained with `pKrKNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference.. 22
8 Histograms for (a) transverse momentum and (b) difference in azimuthal angle of Z^0-pair after full shower for `pKrkNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` and showered leading order as reference. Logarithmic binning is employed to emphasize hadronisation region. 23
9 Histograms for the leading jet transverse momentum distribution after first emission obtained with `MAcNLOPS`, with `MC@NLO` as reference. 24
10 Histograms of only $\mathbb{H}$-event and positive $\mathbb{H}$-event contribution to the leading jet transverse momentum distribution for two different $p_T^{j_1}$ ranges. Negative bins are reflected and plotted in dotted lines. 25

11 Histograms showing the separate effects of removing negative $\mathbb{H}$-events and `MAcNLOPS` vetoing, with `MAcNLOPS` and its input `MC@NLO` for reference. 25
12 Histograms for distribution of difference in azimuthal angle of Z^0-pair after (a) the first emission and (b) full shower for `MAcNLOPS`, with `MC@NLO` as reference. . . 26
13 Histograms for the leading jet rapidity distribution after first emission obtained with `MAcNLOPS`, with `MC@NLO` as reference. 27
14 Histograms for (a,b) invariant mass and (c,d) rapidity separation of the Z^0-pair after first emission (left) and full shower (right) obtained with `MAcNLOPS`, with `MC@NLO` as reference. 27
15 Histograms for (a) transverse momentum and (b) difference in azimuthal angle of Z^0-pair after full shower for `MAcNLOPS`, with `MC@NLO` and showered leading order as reference. Logarithmic binning is employed to emphasize hadronisation region. 28
16 Plots of $\overline{\text{MS}}$ and Krk PDFs at scales 3 GeV (top) and 100 GeV (bottom). 33

1 Introduction

In modern high-energy physics, event generators have become essential, with wide usage on the theoretical side for making quantitative predictions and testing models, as well as on the experimental side for event reconstruction from real data collected at collider experiments. A great effort is thus put into development of event generators, such as PYTHIA [1], HERWIG [2] and SHERPA [3].

In the past decades, event generators capable of producing results at next-to-leading order (NLO) precision in the strong coupling constant α_S have become state of the art in this field. This implies interfacing *hard-event generation* according to the NLO distribution with a *parton-shower* algorithm that resums the large logarithms by performing cascading emissions. The construction of such an interface requires careful physical considerations, as overlaps in contributions can arise. The interfaces addressing this problem are known as *NLO matching schemes*.

The first NLO matching scheme was `MC@NLO` [4], which succeeds in the task but has the inherent property of allowing negative weights, which is to say that while all events have weights uniform in magnitude, some weights will be negative. The presence of negative weights in turn leads to a lowered precision, which can only be remedied by generating a larger event sample. However, as subsequent steps, detector simulation in particular, is significantly more demanding in computational resources, a larger event sample poses an efficiency problem.

To address this problem, several new matching schemes were proposed in the following years, including `POWHEG` [5], `KrkNLO` [6] and `MAcNLOPS` [7], all of which are free of negative weights. While being the earliest and most developed, `POWHEG` is not straightforward to extend to matching with next-to-leading logarithmic (NLL) shower[1] [8], which is not the case for `KrkNLO` and `MAcNLOPS`. Moreover, `KrkNLO` and `MAcNLOPS` are more akin to `MC@NLO` as all three require the explicit reconstruction of the shower kernel used for parton shower.

Among the two, `KrkNLO` has been tested on various colour-singlet processes [9, 10] with an implementation in HERWIG7 [2]. In this implementation, both the virtual and real emission corrections are included as a multiplicative reweight, imposing the requirement of using PDFs transformed to the Krk (MC) factorisation scheme, which is presented fully in Ref. [11]. Upon further inspection built on Ref. [7], it was realised that this requirement can be relaxed if the implementation is modified so that the virtual corrections are included additively while the real emission contribution continues to be included multiplicatively. This is formally equivalent to original `KrkNLO`, but allows usage of the standard $\overline{\text{MS}}$ [12] factorisation scheme.

On the other hand, `MAcNLOPS` has not been implemented to date. Moreover, in the context of NLO matching schemes, `MAcNLOPS` arguably has more desirable features in production of unweighted and positive-definite event samples, along with general applicability.

The main subject of study for this thesis will be the implementation of `KrkNLO` and `MAcNLOPS` using MADGRAPH5_AMC@NLO [13] for $pp \to Z^0 Z^0$, an important process at the Large Hadron Collider (LHC). For `KrkNLO`, we will use the aforementioned modified implementation, giving insight into the impact of factorisation scheme choice. Whereas for `MAcNLOPS`, this will

[1] It is not impossible, as illustrated in Ref. [8], but doing so will require additional care to not spoil NLL accuracy.

be the first study of its practical performance.

To present the efforts and results, the thesis opens with an outline of the concepts relevant to hard event generation and parton shower in Sect. 2, laying foundation for the discussion of their interface, which is done in Sect. 3 with dedicated subsections for `MC@NLO`, `KrkNLO` and `MAcNLOPS`. Then, the specifics of the implementation will be presented in Sect. 4. The results with analyses are in Sect. 5, followed by a conclusion in Sect. 6 that discusses the implications.

2 Background

To simulate entire collision events, modern event generators typically make an artificial stage division. Among the stages, the first two stand in the perturbative regime, generating the hard scattering event and the cascading emissions that lowers the scale to the hadronisation scale $t_{\text{cut}} \sim 1$ GeV. This split is motivated by the differing techniques used for simulation, which comes from both physical and algorithmic efficiency considerations.

In this section, we will present formalisms for hard event generation and parton shower, enabling the discussion towards their interface, which is the main object of study for this thesis. In addition, we briefly discuss phase space mappings, highlighting a subtlety that will be addressed.

2.1 Fixed-Order NLO

Initiating the simulation is the generation of a set of events associated with a hard scale. This is done by sampling a distribution composed of active contributions at some fixed-order. For modern event generators, the precision is typically taken to be next-to-leading order (NLO).

At NLO, one must consider, beyond the leading-order (LO) contributions, virtual corrections from one-loop diagrams and real contributions from one-emission diagrams. Therefore, after including these terms, the differential cross-section $d\sigma_{\text{NLO}}$ should be

$$\begin{aligned} d\sigma_{\text{NLO}} = \mathcal{B}(\Phi_B)d\Phi_B + &\left(\mathcal{V}(\Phi_B) + \int \mathcal{I}^{\text{FKS}}(\Phi_{\text{limit}})d\Phi_R\right)d\Phi_B + \\ &+ \mathcal{C}_+^{\text{FS}}\big(z_+, \Phi_B^+; \mu_F\big)dz_+ d\Phi_B^+ + \mathcal{C}_-^{\text{FS}}\big(z_-, \Phi_B^-; \mu_F\big)dz_- d\Phi_B^- + \\ &+ \big[\mathcal{R}(\Phi) - \mathcal{I}^{\text{FKS}}(\Phi_{\text{limit}})\big]d\Phi, \end{aligned} \tag{2.1}$$

where Φ_B and Φ parametrise the phase space of the n-body and $n+1$-body processes respectively. Furthermore, it is assumed that the phase space mapping is constructed such that $d\Phi = d\Phi_B d\Phi_R$, where Φ_R parametrises the extra emission. The physical contributions are denoted by $\mathcal{B},\ \mathcal{V},\ \mathcal{R}$, representing the LO (Born) contribution, virtual correction and real contribution. By the Kinoshita-Lee-Nauenberg theorem [14, 15], $\mathcal{V}$ and $\mathcal{R}$ contain cancelling infrared (IR) singularities. To achieve separate convergence, the Frixione-Kunszt-Signer (FKS) subtraction terms $\mathcal{I}^{\text{FKS}}$ are inserted to remove singularities in both terms respectively [16]. Formally, they rely on the idea of a plus-prescription and are therefore not evaluated at the same phase space point, but rather at a singular limit. This is noted with the usage Φ_{limit} instead of Φ. Finally, $\mathcal{C}_\pm^{\text{FS}}$ are the factorisation scheme dependent collinear remnant terms, with $z_\pm$ being rescaling factors on the momenta of the $\pm$-rapidity initial-state partons respectively, $\Phi_B^{\pm}$ the configuration with such rescaled momentum and μ_F the factorisation scale. For brevity, both $\mathcal{I}^{\text{FKS}}$ and $\mathcal{C}_{\pm}^{\text{FS}}$ will be suppressed in the remainder of this thesis unless explicitly relevant.

In practice, there is often more than one subprocess contributing to the total NLO cross-section. It is then more accurate to write

$$\mathcal{B}(\Phi_B) = \sum_b \mathcal{B}^{(b)}(\Phi_B) \tag{2.2}$$

$$\mathcal{V}(\Phi_B) = \sum_b \mathcal{V}^{(b)}(\Phi_B) \tag{2.3}$$

$$\mathcal{R}(\Phi_R) = \sum_r \mathcal{R}^{(r)}(\Phi_R), \tag{2.4}$$

where $\{b\}$ and $\{r\}$ represents the sets of Born and real subprocesses. However, for brevity, this sum of channels is often suppressed in expressions.

Due to the different phase space dependence of $\mathcal{I}^{\text{FKS}}$, an attempt to decompose Eq.(2.1) into $(\mathcal{B}+\mathcal{V}+\mathcal{I}^{\text{FKS}})$ and $(\mathcal{R}-\mathcal{I}^{\text{FKS}})$ will fail due to inconsistent phase space dependence. Therefore, fixed-order NLO can only generate weighted events. Such an event sample will have a lower precision compared to an event sample of the same size containing events with uniform weights.

2.2 Parton Showers

For a given hard event, a parton shower algorithm generates cascading emissions that describe the evolution from the input hard scale to the hadronisation scale. Generally, these emissions are classified into initial-state radiations (ISRs) and final-state radiations (FSRs), arising from incoming and outgoing particles in the hard event respectively. As this thesis studies the process $p\,p \to Z^0\,Z^0$, one should be concerned only with ISRs since the Z^0-pair cannot emit. However, we will start with a brief discussion of FSRs as they best illustrate the idea of a parton shower.

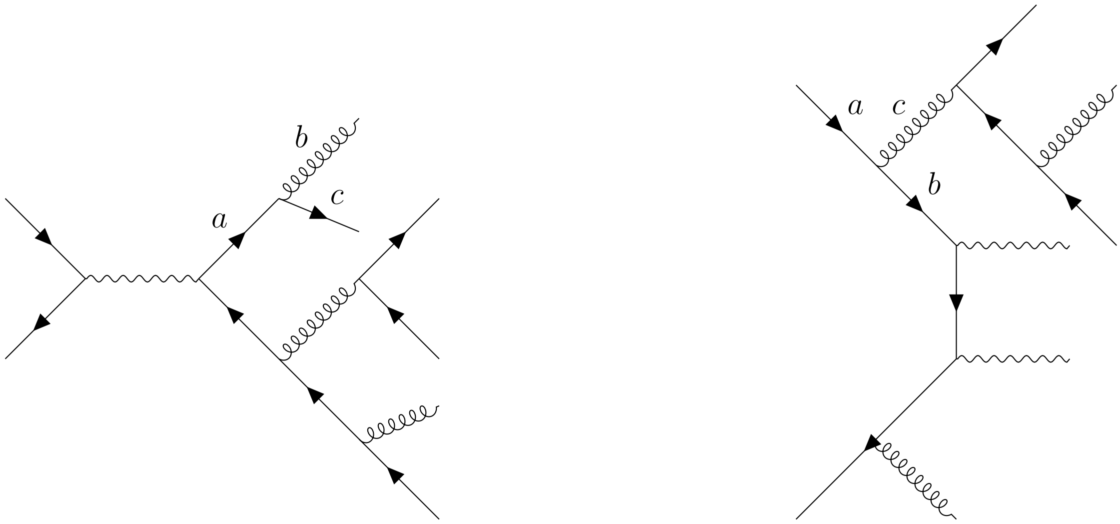
For FSRs (illustrated in Fig. 1a), the emission probability of parton a is given by

$$d\mathcal{P}_a^{\text{FSR}}(t) = \frac{dt}{t}\frac{\alpha_S}{2\pi}\sum_{b,c}\int P_{a\to bc}(z)dz, \tag{2.5}$$

where z is an energy fraction measure such that one branch has energy fraction z and the other $(1-z)$; t is an ordering variable that differs based on the type of shower employed (e.g. transverse momentum p_T^2 or virtuality Q^2); and $P_{a\to bc}$ is a DGLAP splitting kernel. It should be noted that often $z \in [z_{\text{min}}, z_{\text{max}}]$ due to kinematical constraints.

For ISRs (illustrated in Fig. 1b), a backwards evolution perspective is taken, ensuring that the partons participating in the hard scattering can be recreated. This means that while the branching is still $a \to bc$, the initial-state parton is taken to be particle b and one looks for the probability of such parton being a product of a branching.

The additional complication arises from the parton distribution functions (PDFs) that must be taken into consideration as initial-state partons come from the colliding protons. To this end,



(a) (b)

Figure 1: Illustrations of (a) FSR and (b) ISR cascades

the emission probability must be modified to include a ratio of PDFs before and after emission

$$d\mathcal{P}_b^{\text{ISR}}(t) = \frac{dt}{t}\frac{\alpha_S}{2\pi}\sum_a \int P_{a\to bc}(z)\frac{x' f_a(x',t)}{x f_b(x,t)}dz, \tag{2.6}$$

where f is a PDF and $x',\ x$ are Bjorken x's. Note that for ISR z is typically taken to represent the energy fraction of parton b, leading to $z = x/x'$.

With emission probabilities listed, a related quantity in the Sudakov factor [17] $\Delta(t_1, t_2)$, representing the probability to have no emission between t_1 and t_2, can be defined as

$$\Delta(t_1,t_2) = \prod_{t_2,t_2+dt,\ldots,t_1}\Big(1 - d\mathcal{P}(t)\Big) = \exp\Big(-\int_{t_2}^{t_1} d\mathcal{P}(t)\Big), \tag{2.7}$$

where it assumed that $t_2 < t_1$ and subscripts are dropped for brevity.

Then, generating the first shower emission is equivalent to sampling

$$\mathcal{P}_1(t) = \Delta(t_i,t)\delta(t - t_{\text{cut}}) + \frac{d\mathcal{P}(t)}{dt}\Delta(t_i,t), \tag{2.8}$$

where t_i is an input parameter setting the maximum scale the shower can emit at, often referred to as the *shower starting scale* and t_{cut} is the IR cut-off at hadronisation scale. The first term corresponds to the probability of having no shower emission at all, while the second is a conditional probability given as the product of emission probability at scale t and probability of having no emission at higher scale $t < t' < t_i$.

The shower unitarity condition can be trivially derived as the second term is a total derivative, giving

$$\int_{t_{\text{cut}}}^{t_i} \mathcal{P}_1(t)dt = \Delta(t_i,t_{\text{cut}}) + \int_{\Delta(t_i,t_{\text{cut}})}^{1} d\Delta(t_i,t) = 1. \tag{2.9}$$

This property ensures that the shower does not modify the total cross-section but does redistribute the events.

2.3 Phase Space Mappings

A crucial part of both the NLO computation and parton shower is the construction of real phase space point Φ from a given Born phase space point Φ_B and a set of variables parametrising the emission, i.e. the mapping $\mathcal{F} : \{\Phi_B, u, v, w\} \to \Phi$, where $\{u, v, w\}$ are placeholders for the aforementioned variables. Such a mapping is not unique as the choices of recoil scheme and variables are largely arbitrary as long as four-momentum conservation and on-shell conditions are respected for the final configuration.

In this thesis, we consider the FKS mapping $\mathcal{F}^{\text{FKS}} : \{\Phi_B, \xi, y, \phi\} \to \Phi$ and shower mapping $\mathcal{F}^{\text{PS}} : \{\Phi_B, z, t, \phi\} \to \Phi$, employed by hard event generation and parton shower respectively. Note that $\{\xi, y, \phi\}$ are the FKS variables related to the energy, collinearity and azimuthal angle of the emission [16], while $\{z, t, \phi\}$ are the shower variables for energy fraction of the the emission, shower scale it is emitted at and again the azimuthal angle [1].

In the partonic centre-of-mass (CM) frame, FKS mapping creates the four-momentum of the emission as

$$p_{\text{em}} = \frac{\sqrt{\hat{s}}}{2}\xi\big(1,\ \cos\phi\sqrt{1-y^2},\ \sin\phi\sqrt{1-y^2},\ y\big). \tag{2.10}$$

To conserve four-momentum, all final-state particles (except the emission) are boosted along the direction $\boldsymbol{x} = -(\cos\phi, \sin\phi, 0)$, which allows the initial-state partons to remain collinear with the beam-axis with CM energy $\hat{s} = \hat{s}_B/(1-\xi)$. This is an example of a global recoil scheme.

On the other hand, the specific form of the emission four-momentum in the shower mapping depends on the shower variables used, which can differ for different choices. We consider specifically PYTHIA8's choice for its default shower mode, where the variables are chosen to be $\{z, p_T^2, \phi\}$. Similar to the preceding section, the branching $a \to bc$ is considered as ISR, where b is the parton that would participate in the hard scattering. In the partonic CM frame, the branches have four-momenta [1]

$$p_b = \left(\frac{\sqrt{\hat{s}}}{2}z,\ p'_T\cos\phi,\ p'_T\sin\phi,\ \pm\frac{\sqrt{\hat{s}}}{2}\left(z + \frac{2Q^2}{\hat{s}}\right)\right) \tag{2.11}$$

$$p_c = \left(\frac{\sqrt{\hat{s}}}{2}(1-z),\ -p'_T\cos\phi,\ -p'_T\sin\phi,\ \pm\frac{\sqrt{\hat{s}}}{2}\left(1 - z - \frac{2Q^2}{\hat{s}}\right)\right), \tag{2.12}$$

where $p_T'^2 = p_T^2 - Q^4/\hat{s}$ and $Q^2 = -m_b^2$.

As the initial-state partons need to remain collinear with the beam-axis, the final-state partons must recoil against the emission to conserve four-momentum for the first emission. This is also a global recoil scheme, but the specifics will generally differ compared to the FKS mapping. For subsequent emissions, a local recoil scheme is also possible, where only one particle recoils. For more details on this matter, we refer the interested reader to Ref. [1].

Note that with such a construction, it is ensured that the invariant mass of the system $(b + s)$, where s is the other initial-state parton, is unchanged. Noting that in the partonic CM frame

$p_s = \sqrt{\hat{s}}/2(1, 0, 0, \mp 1)$, one can obtain

$$(p_b + p_s)^2 = -Q^2 + 2p_b \cdot p_s = -Q^2 + \frac{\hat{s}}{2}\left(z + z + \frac{2Q^2}{\hat{s}}\right) = z\hat{s}. \tag{2.13}$$

Since $z = x_b/x_a$, $z\hat{s}$ is exactly the invariant mass of the $(b + s)$ system prior to generating this branching history. This achieves the goal of a backwards evolution, where the original hard event is successfully reconstructed.

It should also be remarked that the ordering variable p_T^2 does not correspond exactly to the transverse momentum of the emission, but is instead given by the relation $p_T^2 = (1 - z)Q^2$. However, it has the advantage of being linear with Q^2, which is desirable for multiple ordered emissions. This was not a concern for FKS mapping as it is used mainly to treat first emissions.

With the FKS and shower mappings defined, we can also define the corresponding inverse mappings

$$\mathcal{F}_{\text{FKS}}^{-1} : \Phi \to \{\Phi_B, \xi, y, \phi\} \text{ and } \mathcal{F}_{\text{PS}}^{-1} : \Phi \to \{\Phi_B, z, t, \phi\}.$$

We discuss in particular the inverse FKS mapping, where the FKS variables can be reconstructed as

$$\xi = \frac{p_{\text{em}}^0}{\sqrt{\hat{s}}/2},\ y = \frac{p_{\text{em}}^3}{p_{\text{em}}^0},\ \phi = \arctan\left(\frac{p_{\text{em}}^2}{p_{\text{em}}^1}\right). \tag{2.14}$$

The partonic CM energy for the Born subprocess is $\hat{s}_B = \hat{s}(1 - \xi)$ and the other final-state particles are subject to an inverse transverse boost along the direction $\boldsymbol{x} = (\cos\phi, \sin\phi, 0)$.

Given the arbitrariness of the recoil scheme, equivalence of the mappings are true only on a variable level. To see this, we consider a parton shower mapping, giving

$$\mathcal{F}^{\text{PS}}(\Phi_B, z, t, \phi) = \Phi^{\text{PS}}.$$

Then, let there be an operation $\mathcal{C}^{\text{FKS}} : \Phi \to \{\xi, y, \phi\}$ that extracts FKS variables from a given real phase space point Φ. Applying this operation gives

$$\mathcal{C}^{\text{FKS}}\left(\Phi^{\text{PS}}\right) = \{\xi, y, \phi\}.$$

Now generally, an FKS mapping using these variables and the Born phase space point Φ_B does not reach the same real phase space point Φ. That is,

$$\mathcal{F}^{\text{FKS}}(\Phi_B, \xi, y, \phi) = \Phi^{\text{FKS}} \neq \Phi^{\text{PS}},$$

despite agreeing on the variable level. Conversely for

$$\mathcal{F}_{\text{FKS}}^{-1}\left(\Phi^{\text{PS}}\right) = \{\Phi'_B, \xi, y, \phi\},$$

we always have $\{\xi, y, \phi\} = \mathcal{C}^{\text{FKS}}\left(\Phi^{\text{PS}}\right)$ but generally $\Phi'_B \neq \Phi_B$, meaning that the momenta for the particles in the Born subprocess will be different. The same argument holds with the exchange PS $\leftrightarrow$ FKS everywhere.

3 NLO Matching

With both NLO hard event generation and parton shower in place, one should now turn to their interface. Such an interface, known as an *NLO matching scheme*, is needed since directly feeding an event sample generated from fixed-order NLO to the parton shower algorithm gives a distribution that is not NLO-accurate.

This section will start by demonstrating how this problem arises, followed by discussions of three NLO matching schemes in `MC@NLO` [4], `KrkNLO` [6] and `MAcNLOPS` [7].

3.1 No Matching

By definition, an observable O is NLO-accurate if

$$\langle O\rangle_{\text{NLO}} = \int \big\{\mathcal{B}(\Phi_B) + \mathcal{V}(\Phi_B)\big\} O(\Phi_B) d\Phi_B + \int \mathcal{R}(\Phi) O(\Phi) d\Phi + \mathcal{O}(\text{NNLO}). \tag{3.1}$$

Therefore, with fixed-order NLO interfaced with parton shower, we should have

$$\langle O\rangle_{\text{NLO+PS}} = \langle O\rangle_{\text{NLO}}, \tag{3.2}$$

up to higher order terms. This requirement must be satisfied by all valid NLO matching schemes and is thus known as the *NLO matching condition*. In particular, this needs only be checked after the first shower emission since all subsequent emissions are beyond the desired accuracy.

For the moment, we consider $d\sigma_{\text{NLO}}$, given in Eq.(2.1), redistributed by $d\mathcal{P}_1$, given in Eq.(2.8). This corresponds to having no additional matching in place, where the events with an n-body configuration in fixed-order results obtained from sampling $d\sigma_{\text{NLO}}$ are showered[2]. To this end, we remark that upon rewrites, the distribution sampled by the first shower emission is

$$\frac{d\mathcal{P}_1(t,\Phi_B)}{d\Phi_R} = \Delta(t_i,t,\Phi_B)\delta\big(t(\Phi)-t_{\text{cut}}\big) + \frac{\mathcal{K}(\Phi)}{\mathcal{B}(\Phi_B)}\Delta(t_i,t,\Phi_B)\Theta\big(t_{\text{cut}} < t(\Phi) < t_i(\Phi)\big), \tag{3.3}$$

where we have made the substitution $d\mathcal{P}/dt = \mathcal{K}/\mathcal{B}$ and consequently

$$\Delta(t_1,t_2,\Phi_B) = \exp\left(-\int_{t_{\text{cut}}<t_2(\Phi)<t_1(\Phi)} \frac{\mathcal{K}(\Phi)}{\mathcal{B}(\Phi_B)} d\Phi\right), \tag{3.4}$$

where it is conventional to suppress the starting scale if $t_1 = t_i$.

The quantity $\mathcal{K}$ is known as the *shower kernel* and it is implied that $\mathcal{K} = \sum_\alpha \mathcal{K}^{(\alpha)}$, where α denotes different types of possible branching. Note also that $\mathcal{K}$ is an approximation of the real contribution that is exact in the collinear limit.

With the first shower emission, the differential cross-section is redistributed and becomes

$$d\sigma_{\text{NLO+PS}} = d\sigma_{\text{B+V}} d\mathcal{P}_1 + d\sigma_R$$

[2]The events with $n+1$ particles already contain the first emission.

$$= (\mathcal{B}+\mathcal{V})\Delta(t_{\text{cut}},\Phi_B)d\Phi_B + (\mathcal{B}+\mathcal{V})\frac{\mathcal{K}}{\mathcal{B}}\Delta(t(\Phi),\Phi_B)\Theta\big(t(\Phi)-t_{\text{cut}}\big)d\Phi + \mathcal{R}d\Phi,$$

where the delta function is used to eliminate the Φ_R-integral in the first term.

Applying this distribution to an arbitrary observable O gives

$$\begin{aligned}\langle O\rangle_{\text{NLO+PS}} &= \int(\mathcal{B}+\mathcal{V})\Delta(t_{\text{cut}},\Phi_B)O(\Phi_B)d\Phi_B + \\ &\quad + \int(\mathcal{B}+\mathcal{V})\frac{\mathcal{K}}{\mathcal{B}}\Delta(t(\Phi),\Phi_B)\Theta\big(t(\Phi)-t_{\text{cut}}\big)O(\Phi)d\Phi + \int\mathcal{R}O(\Phi)d\Phi \\ &= \int(\mathcal{B}+\mathcal{V})O(\Phi_B)\left[\Delta(t_{\text{cut}},\Phi_B) + \int_{t(\Phi)>t_{\text{cut}}}\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)d\Phi_R\right]d\Phi_B + \\ &\quad + \int_{t(\Phi)>t_{\text{cut}}}(\mathcal{B}+\mathcal{V})\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)\big(O(\Phi)-O(\Phi_B)\big)d\Phi + \int\mathcal{R}O(\Phi)d\Phi \\ &= \int(\mathcal{B}+\mathcal{V})O(\Phi_B)d\Phi_B + \int_{t(\Phi)>t_{\text{cut}}}(\mathcal{B}+\mathcal{V})\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)\big(O(\Phi)-O(\Phi_B)\big)d\Phi + \\ &\quad + \int\mathcal{R}O(\Phi)d\Phi,\end{aligned}$$

where arguments are suppressed when unambiguous for brevity. Note also that the term in the square brackets simplifies to one in the second last step due to shower unitarity.

To further simplify this expression, it should be noted that for non-singular quantities, one can safely remove terms beyond NLO. In the second term, since $\big(O(\Phi)-O(\Phi_B)\big)$ provides the suppression for the singularities in $\mathcal{K}$, it can be simplified into

$$\int_{t(\Phi)>t_{\text{cut}}}\mathcal{K}\big(O(\Phi)-O(\Phi_B)\big)d\Phi + \mathcal{O}(\text{NNLO}),$$

as $\mathcal{V}\sim\mathcal{O}(\alpha_S)$ and $\Delta = 1+\mathcal{O}(\alpha_S)$, which, when multiplied with $\mathcal{K}\sim\mathcal{O}(\alpha_S)$, produces higher order terms.

Substituting back into the original expression gives, up to higher order corrections

$$\langle O\rangle_{\text{NLO+PS}} = \int(\mathcal{B}+\mathcal{V})O(\Phi_B)d\Phi_B + \int_{t(\Phi)>t_{\text{cut}}}\mathcal{K}\big(O(\Phi)-O(\Phi_B)\big)d\Phi + \int\mathcal{R}O(\Phi)d\Phi, \quad (3.5)$$

which differs from Eq.(3.1) by the $\mathcal{O}(\alpha_S)$ middle term and is thus not NLO-accurate.

Heuristically, the discrepancy between Eq.(3.1) and Eq.(3.5) should be attributed to the fact that the same Feynman diagram can arise from considering a real subprocess (when doing NLO event generation) as well as a Born subprocess with a shower emission. Therefore, this problem is often referred to as *double-counting*.

3.2 MC@NLO

As the earliest NLO matching scheme proposed, `MC@NLO` [4] relies on the idea of removing the double-counted contributions when generating hard events. This contribution can be easily identified by comparing Eq.(3.1) with Eq.(3.5), giving the differential cross-section for MC@NLO as

$$d\sigma_{\text{MC@NLO}} = \mathcal{B}(\Phi_B)d\Phi_B + \left(\mathcal{V}(\Phi_B) + \int \mathcal{K}(\Phi)d\Phi_R\right)d\Phi_B + \big(\mathcal{R}(\Phi) - \mathcal{K}(\Phi)\big)d\Phi, \tag{3.6}$$

where $\mathcal{I}^{\text{FKS}}$ is absent since $\mathcal{K}$ is responsible for cancelling the singularities[3].

Same as before, to check NLO matching condition, we consider the differential cross-section after first shower emission. In the case of MC@NLO, this will be

$$d\sigma_{\text{MC@NLO+PS}} = \left(\mathcal{B}+\mathcal{V}+\int \mathcal{K}d\Phi_R\right)\Delta(t_{\text{cut}},\Phi_B)+\left(\mathcal{B}+\mathcal{V}+\int \mathcal{K}d\Phi_R\right)\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)d\Phi+(\mathcal{R}-\mathcal{K})d\Phi, \tag{3.7}$$

where the Heaviside function is implicit.

The derivation is then effectively unmodified, except the replacements $\mathcal{V} \to \mathcal{V} + \int \mathcal{K}d\Phi_R$ and $\mathcal{R} \to \mathcal{R} - \mathcal{K}$. Therefore, instead of Eq.(3.5), we arrive at

$$\begin{aligned}
\langle O\rangle_{\text{MC@NLO+PS}} &= \int (\mathcal{B}+\mathcal{V})O(\Phi_B)d\Phi_B + \int \mathcal{K}O(\Phi_B)d\Phi_B d\Phi_R + \int \mathcal{K}\big(O(\Phi) - O(\Phi_B)\big)d\Phi + \\
&\quad + \int (\mathcal{R}-\mathcal{K})O(\Phi)d\Phi \\
&= \int \big\{\mathcal{B}(\Phi_B) + \mathcal{V}(\Phi_B)\big\}O(\Phi_B)d\Phi_B + \int \mathcal{R}(\Phi)O(\Phi)d\Phi + \mathcal{O}(\text{NNLO}) \\
&= \langle O\rangle_{\text{NLO}},
\end{aligned}$$

demonstrating that MC@NLO satisfies the NLO matching condition.

Crucially, since $\mathcal{K}$ has the same phase space dependence as $\mathcal{R}$, generation of unweighted events is now possible. To this end, we consider two classes of events corresponding to subprocesses with n and $n+1$ particles, known as $\mathbb{S}$- and $\mathbb{H}$-events respectively. One can define separate generating functionals for them as

$$\mathcal{F}^{(\mathbb{S})}(\Phi_B) = \mathcal{B}(\Phi_B) + \mathcal{V}(\Phi_B) + \int \mathcal{K}(\Phi)d\Phi_R \equiv \widetilde{\mathcal{B}}(\Phi_B) \tag{3.8}$$

$$\mathcal{F}^{(\mathbb{H})}(\Phi) = \mathcal{R}(\Phi) - \mathcal{K}(\Phi). \tag{3.9}$$

A normalised probability distribution can then be constructed at every accepted phase space point, providing a randomised decision for which subprocess to output via cumulative sampling. Each event generated this way will be associated with a uniform (up to a sign) weight, representing an increased precision compared to a weighted event sample of the same size.

Moreover, the shower starting scales are set in MC@NLO such that the correct distribution is reproduced at low- and high-p_T regions. An example is shown in Fig. 2. From the ratio with

[3]This is true for $pp \to Z^0Z^0$, where $\mathcal{R} \to \mathcal{K}$ in both soft and collinear limits. In the general case, $\mathcal{K}$ must be replaced by $\mathcal{I}^{\text{FKS}}$ in the various limits. This replacement is governed by two G-functions with arguments $(1-\xi)$ and y respectively, where a G-function is a generalised step-function with limits $\lim_{|x|\to 1}\mathcal{G}(x) = 0$ and $\lim_{x\to 0}\mathcal{G}(x) = 1$ [18].

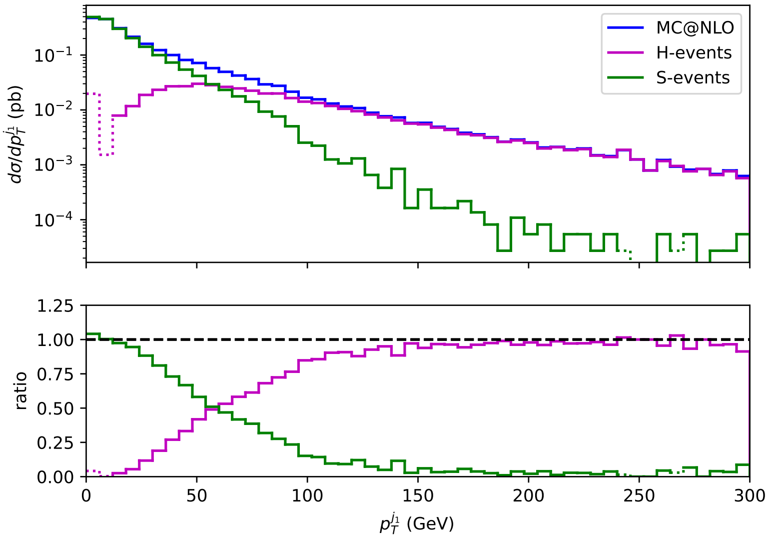


Figure 2: Example of a histogram for transverse momentum of the leading jet, with contributions from $\mathbb{S}$- and $\mathbb{H}$-events shown separately. Negative bins are reflected and labelled using dotted lines.

full `MC@NLO`, one observes the $\mathbb{S}$-events dominating the low-p_T region, followed by a smooth transition towards the $\mathbb{H}$-dominated high-p_T region. The shower starting scale dictates the dead zone implemented in $\mathcal{K}$, setting it to zero at phase space points corresponding to an emission at higher scale. This means that $\mathcal{F}^{(\mathbb{H})} \to \mathcal{R}$ in high-p_T regions, reproducing the correct NLO distribution. In the low-p_T region, $\mathcal{R} \approx \mathcal{K}$, eliminating the $\mathbb{H}$-event contribution, producing a distribution that matches in shape with that given by showering a leading-order event sample.

However, an important drawback of `MC@NLO` lies in the fact that both Eqs.(3.8) and (3.9) are not positive definite. This gives events that are associated with a negative weight, whose presence leads to an increase in imprecision. To achieve the same level of precision as a positive-definite event sample, a mixed event sample with fraction f being negative must be larger by a factor of $1/(1-2f)^2$ [19]. Reduction of negative weight fraction is thus very desirable to reduce the already increasing computational burden.

To this end, an analysis on the origin of negative weights should be done. For $\mathbb{S}$-events, while negative weights can arise from the virtual contribution being locally negative with large magnitude, this effect is limited as $\mathcal{V} \sim \mathcal{O}(\alpha_S)$ with singularities subtracted. The main source of negative $\mathbb{S}$-weights arises when $\int \mathcal{K} d\Phi_R$ cannot be done analytically. This requires a re-insertion of $\mathcal{I}^{\text{FKS}}$, giving $\left(\mathcal{B}+\mathcal{V}+\mathcal{I}^{\text{FKS}}\right)$ and $\int \left(\mathcal{K}-\mathcal{I}^{\text{FKS}}\right) d\Phi_R$ as two components of $\widetilde{\mathcal{B}}$. The latter, being computed numerically is now subject to imprecisions[4]. Such imprecision leads to locally large

[4]In a typical implementation, one samples $\mathbb{S}$-events in the $(n+1)$-body phase space as well, meaning that one takes the approximation $\int \left(\mathcal{K} - \mathcal{I}^{\text{FKS}}\right) d\Phi_R = \mathcal{K}(\Phi) - \mathcal{I}^{\text{FKS}}(\Phi)$, where Φ is the sampled point. This is allowed since a great number of events are sampled in total. But for each event, the evaluation of the integral will be very

and negative contributions. As such, complete removal of negative $\mathbb{S}$-weights (up to the allowed limit set by possibly negative $\mathcal{V}$) can be achieved at the expense of large computational resource consumption to evaluate the Φ_R-integral precisely. This is the idea behind the folding method [20]. Other more efficient methods of varying effectiveness are also proposed over the years (e.g. [21–23]). In this thesis, it is then reasonable to neglect these negative weights and assume $\widetilde{\mathcal{B}}$ to be positive-definite.

On the other hand, little can be done for negative $\mathbb{H}$-events as it arises by construction. Namely, there cannot be a requirement that $\mathcal{K} < \mathcal{R},\ \forall \Phi$. Therefore, formally different matching schemes must be devised to reduce the fraction of negative $\mathbb{H}$-events (e.g. `MC@NLO-`Δ [24]).

3.3 KrkNLO

The more recently proposed `KrkNLO` [6, 9, 10] belongs to a different class of matching schemes that uses a multiplicative reweight to achieve NLO accuracy. In practice, this requires integration of the first shower emission into the hard event generation stage. Depending on the outcome of the first shower emission, different multiplicative reweighting factors are used. To illustrate this, the differential cross-section can be written as

$$d\sigma_{\text{KrkNLO}} = \mathcal{B}(\Phi_B)\Delta^{(m)}(t_{\text{cut}}, \Phi_B)\frac{\widetilde{\mathcal{B}}(\Phi_B)}{\mathcal{B}(\Phi_B)}d\Phi_B + \mathcal{B}(\Phi_B)\frac{\mathcal{K}(\Phi)}{\mathcal{B}(\Phi_B)}\Delta^{(m)}(t, \Phi_B)\frac{\widetilde{\mathcal{B}}(\Phi_B)}{\mathcal{B}(\Phi_B)}\frac{\mathcal{R}(\Phi)}{\mathcal{K}(\Phi)}d\Phi. \tag{3.10}$$

With Born events sampled according to $\mathcal{B}$ showered, the first term corresponds to events with no shower emission until t_{cut} and the second term corresponds to events with a shower emission at scale t. The two reweight functions are then

$$\mathcal{W}_0(\Phi_B) = \frac{\widetilde{\mathcal{B}}(\Phi_B)}{\mathcal{B}(\Phi_B)} \tag{3.11}$$

$$\mathcal{W}_1(\Phi) = \frac{\widetilde{\mathcal{B}}(\Phi_B)}{\mathcal{B}(\Phi_B)}\frac{\mathcal{R}(\Phi)}{\mathcal{K}(\Phi)}. \tag{3.12}$$

Note that as $\mathcal{R}$ enters only as a reweight, the shower must be responsible of generating the hard emissions as well. This is achieved by always setting the shower starting scale to be t_{max}, the maximum allowed by the process. This is denoted by the superscript (m) in Eq.(3.10). Such an option is known as a *power shower*.

To demonstrate that the NLO matching condition is satisfied, we rewrite the second term as

$$\begin{aligned}\widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\frac{\mathcal{R}}{\mathcal{K}} &= \widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\left(\frac{\mathcal{R}}{\mathcal{K}} - 1\right) + \widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\\ &= \frac{\widetilde{\mathcal{B}}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\big(\mathcal{R} - \mathcal{K}\big) + \widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B).\end{aligned}$$

imprecise.

As $\mathcal{R} \to \mathcal{K}$ in the singular limits, the higher order terms can again be dropped in the first term, giving

$$d\sigma_{\text{KrkNLO}} = \widetilde{\mathcal{B}}\Delta^{(m)}(t_{\text{cut}}, \Phi_B)d\Phi_B + \widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)d\Phi + (\mathcal{R} - \mathcal{K})d\Phi + \mathcal{O}(\text{NNLO}),$$

which is exactly $d\sigma_{\text{MC@NLO+PS}}$ given in Eq.(3.7). Therefore, as `MC@NLO` satisfies the NLO matching condition, so does `KrkNLO`.

`KrkNLO` has several advantages. Since the reweight functions are both positive-definite, the produced event sample will be positive-definite. With singularities subtracted for $\mathcal{V}$ and divided out for $\mathcal{R}$, although events are weighted, the unweighting efficiency should be reasonable.

However, returning to the NLO case relevant to us, as $\widetilde{\mathcal{B}}$ takes the form

$$\widetilde{\mathcal{B}}(\Phi_B) = \mathcal{B}(\Phi_B) + \mathcal{V}(\Phi_B) + \int_0^1 \mathcal{C}_+^{\text{FS}}\big(z_+, \Phi_B^+; \mu_F\big)dz_+ + \int_0^1 \mathcal{C}_-^{\text{FS}}\big(z_-, \Phi_B^-; \mu_F\big)dz_-,$$

to use `KrkNLO` implies computing the integrals for the collinear remnants at each Φ_B sampled, which is impossible.

To address this problem, `KrkNLO` was proposed with a specially-defined Krk factorisation scheme [11]. In the Krk scheme, the collinear remnant terms are reduced to

$$\mathcal{C}_\pm^{\text{Krk}} = \frac{1}{2}\Delta^{\text{Krk}}\delta(1 - z_\pm), \tag{3.13}$$

when the factorisation scale is fixed at $\mu_F^2 = \hat{s}$. Using this defining identity and imposing the usual momentum sum rules, one can derive a transformation from PDFs in the $\overline{\text{MS}}$ [12] factorisation schemes to their counterparts in the Krk scheme. Derivations can be found in [9, 11], which are reproduced in Appendix A of this thesis. For colour-singlet processes initiated by a $q\bar{q}$ pair at leading order, $\widetilde{\mathcal{B}}$ simplifies to

$$\widetilde{\mathcal{B}}(\Phi_B) = \mathcal{B}(\Phi_B) + \mathcal{V}(\Phi_B) + \frac{\alpha_S}{2\pi}\Delta_0^{\text{Krk}}, \tag{3.14}$$

where

$$\Delta_0^{\text{Krk}} = C_F\left(2\pi^2 - \frac{3}{2}\right).$$

In this original formulation, `KrkNLO` must be used in conjunction with the Krk factorisation scheme. However, if we instead sample events passed to the shower according to $\widetilde{\mathcal{B}}$, the collinear remnants need not be computed for the reweight functions, allowing the usage of PDFs in any factorisation schemes. This leads to a mixture of additive and multiplicative matching. Taking this view, the differential cross-section should be written in the form

$$d\sigma_{\text{KrkNLO}} = \widetilde{\mathcal{B}}(\Phi_B)\Delta^{(m)}(t_{\text{cut}}, \Phi_B)d\Phi_B + \widetilde{\mathcal{B}}(\Phi_B)\frac{\mathcal{K}(\Phi)}{\mathcal{B}(\Phi_B)}\Delta^{(m)}(t, \Phi_B)\frac{\mathcal{R}(\Phi)}{\mathcal{K}(\Phi)}d\Phi, \tag{3.15}$$

where $\mathbb{S}$-events are sampled and the ones with a shower emission are further reweighted by $\mathcal{R}/\mathcal{K}$. Therefore, the new reweight functions are

$$\mathcal{W}_0'(\Phi_B) = 1 \tag{3.16}$$

$$\mathcal{W}_1'(\Phi) = \frac{\mathcal{R}(\Phi)}{\mathcal{K}(\Phi)}. \tag{3.17}$$

In essence, this re-formulation trades the conceptual straightforwardness of achieving NLO accuracy via pure multiplicative reweight for the freedom of using PDFs in any factorisation scheme, which is appropriate if one is concerned with only NLO matching.

To distinguish between the two perspectives, we will refer to them as `fKrkNLO` and `pKrkNLO` in this thesis, for either a full or a partial multiplicative reweight used to achieve NLO accuracy. For future reference, we list their respective $d\sigma$ expressions again as

$$d\sigma_{\text{fKrkNLO}} = \mathcal{B}\Delta^{(m)}(t_{\text{cut}}, \Phi_B)\frac{\widetilde{\mathcal{B}}}{\mathcal{B}}d\Phi_B + \mathcal{B}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\frac{\widetilde{\mathcal{B}}}{\mathcal{B}}\frac{\mathcal{R}}{\mathcal{K}}d\Phi \tag{3.18}$$

$$d\sigma_{\text{pKrkNLO}} = \widetilde{\mathcal{B}}\Delta^{(m)}(t_{\text{cut}}, \Phi_B)d\Phi_B + \widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta^{(m)}(t, \Phi_B)\frac{\mathcal{R}}{\mathcal{K}}d\Phi. \tag{3.19}$$

Any mention of just `KrkNLO` then refers to the general idea, rather than any particular perspective.

3.4 MAcNLOPS

From preceding discussions, it becomes clear that `MC@NLO` and `KrkNLO` have compensating advantages and drawbacks. Moreover, on the technical side, both schemes require a reconstruction of the shower kernel $\mathcal{K}$, paving way for a unification. Therefore, one can envisage a new matching scheme that conceptually merges `MC@NLO` and `KrkNLO` to generate unweighted and positive-definite event samples, which implies a smaller sample size needed to reach the required precision (derived in Appendix B). By using `KrkNLO` in the negative-$\mathbb{H}$ regions and `MC@NLO` everywhere else, `MAcNLOPS` [7] achieves this exactly.

Similar to `KrkNLO`, `MAcNLOPS` also requires the complete integration of the first shower emission into hard event generation. Since the introduction of `KrkNLO` is to compensate for negative-$\mathbb{H}$ events, it is natural to take the `pKrkNLO` perspective, where $\mathbb{S}$-event generation proceeds as normal `MC@NLO`. This gives the differential cross-section as a combination of Eq.(3.6) and Eq.(3.19) in the form of

$$\begin{aligned} d\sigma_{\text{MAcNLOPS}} = &\ \widetilde{\mathcal{B}}(\Phi_B)\Delta(t_{\text{cut}}, \Phi_B)d\Phi_B + \\ &+ \widetilde{\mathcal{B}}(\Phi_B)\frac{\mathcal{K}}{\mathcal{B}}\Delta(t, \Phi_B)\left[1 + \frac{\mathcal{R}-\mathcal{K}}{\mathcal{K}}\Theta\big(\mathcal{K}-\mathcal{R}\big)\right]d\Phi + \big(\mathcal{R}-\mathcal{K}\big)\Theta\big(\mathcal{R}-\mathcal{K}\big)d\Phi, \end{aligned} \tag{3.20}$$

where the no-emission term is again present in the unmodified form, but a Heaviside function is responsible for separating the positive and negative $\mathbb{H}$-weight regions, allowing different treatments. A rewrite of the term in the square brackets gives

$$1 + \frac{\mathcal{R}-\mathcal{K}}{\mathcal{K}}\Theta\big(\mathcal{K}-\mathcal{R}\big) = \begin{cases} 1\ , & \text{if } \mathcal{R} \geq \mathcal{K} \\ \mathcal{R}/\mathcal{K}\ , & \text{otherwise} \end{cases}.$$

This shows that in the positive-$\mathbb{H}$ regions, the second and third term in Eq.(3.20) corresponds to $\mathbb{S}$-events with a shower emission and $\mathbb{H}$-events respectively, which is identical with `MC@NLO`. In the negative-$\mathbb{H}$ regions, the third term vanishes, and the second term corresponds to multiplicatively-reweighted $\mathbb{S}$-events, which is identical to pKrkNLO.

To demonstrate formally that MAcNLOPS satisfies the NLO matching condition, we expands the square brackets, giving

$$\widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)+\widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)\frac{\mathcal{R}-\mathcal{K}}{\mathcal{K}}\Theta\big(\mathcal{K}-\mathcal{R}\big).$$

The singularities in the second term are suppressed by the difference $(\mathcal{R}-\mathcal{K})$, allowing it to be simplified to

$$(\mathcal{R}-\mathcal{K})\Theta\big(\mathcal{K}-\mathcal{R}\big).$$

Substituting back into Eq.(3.20) gives exactly $d\sigma_{\text{MC@NLO+PS}}$

$$d\sigma_{\text{MAcNLOPS}}=\underbrace{\widetilde{\mathcal{B}}\Delta(t_{\text{cut}},\Phi_B)d\Phi_B+\widetilde{\mathcal{B}}\frac{\mathcal{K}}{\mathcal{B}}\Delta(t,\Phi_B)+(\mathcal{R}-\mathcal{K})}_{d\sigma_{\text{MC@NLO+PS}}}+\mathcal{O}(\text{NNLO}).$$

Crucially, the quantity in the square brackets in Eq.(3.20) is bounded between $[0,1]$ as it becomes $\mathcal{R}/\mathcal{K}$ only when $\mathcal{R}<\mathcal{K}$. Therefore, instead of interpreting it as a multiplicative reweight, it can be seen as a veto probability. The same, albeit trivial, can be said about the Heaviside function in the third term. This interpretation gives

$$\mathcal{P}^{(\mathbb{S})}_{\text{veto}}(\Phi)=1+\frac{\mathcal{R}(\Phi)-\mathcal{K}(\Phi)}{\mathcal{K}(\Phi)}\Theta\big(\mathcal{K}-\mathcal{R}\big) \tag{3.21}$$

$$\mathcal{P}^{(\mathbb{H})}_{\text{veto}}(\Phi)=\Theta\big(\mathcal{R}(\Phi)-\mathcal{K}(\Phi)\big), \tag{3.22}$$

where $\mathcal{P}^{(\mathbb{S})}_{\text{veto}}$ is applied on $\mathbb{S}$-events with an emission and $\mathcal{P}^{(\mathbb{H})}_{\text{veto}}$ on $\mathbb{H}$-events.

Conceptually, one can see that these two veto procedures are compensating. Consider a histogram of an observable sensitive to singularities, e.g. transverse momentum of the leading jet $p_T^{j_1}\propto\xi\sqrt{1-y^2}$. $\mathcal{P}^{(\mathbb{H})}_{\text{veto}}$ removes the negative $\mathbb{H}$-weights, which, as shown in Fig. 2, will be present in the near-singular regions, leading to an increase of bin heights in the such regions. On the other hand, $\mathcal{P}^{(\mathbb{S})}_{\text{veto}}$ is only active in these regions and decreases as the difference $\big(\mathcal{K}-\mathcal{R}\big)$ grows. Therefore, one can view the veto as being based on a normalised size of said difference. As the regions with a larger difference are also more likely to be sampled as $\mathbb{H}$-events, the increased likelihood of such shower emissions being vetoed implies cancellation of the effect of negative $\mathbb{H}$-event removal.

With this veto probability interpretation, event weights need not be modified and all negative-$\mathbb{H}$ events will be removed, allowing generation of an unweighted and positive-definite event sample.

Possible variants of MAcNLOPS also exist, including one where the formal substitution $\mathcal{K}\to c\mathcal{K}$ with c being some constant is made[5]. This should be viewed as a generalisation of Eq.(3.20). The additional control enabled by the parameter c can prove beneficial in cases. However, as this effect was not explored in this thesis, we will not elaborate on it further and instead direct the interested reader to Sect. 4 of Ref. [7].

[5]Except in the Sudakov factor.

4 Implementation

Among the three matching schemes presented in the preceding section, MC@NLO is by far the most established, with complete and robust implementations in most modern event generators. This is not the case for KrkNLO and MAcNLOPS. The fKrkNLO variant of KrkNLO has been implemented only for a selected set of colour-singlet processes [9, 10] in HERWIG7. MAcNLOPS is yet to be implemented for any process.

In this section, we present the implementation of pKrkNLO, which allows any factorisation scheme to be used. Then, we will explain a minimalistic implementation of MAcNLOPS. As the practical implementation involves MADGRAPH5_AMC@NLO (MG5_AMC) [13] interfaced with the PYTHIA8 [1] shower, this section will end with a discussion on the relevant technical details common to both matching schemes, leaving scheme-specific details to the respective sections.

4.1 pKrkNLO with MG5_AMC

Recall that the differential cross-section given by pKrkNLO is

$$d\sigma_{\text{pKrkNLO}} = \widetilde{\mathcal{B}}(\Phi_B)\Delta^{(m)}(t_{\text{cut}}, \Phi_B)d\Phi_B + \widetilde{\mathcal{B}}(\Phi_B)\frac{\mathcal{K}(\Phi)}{\mathcal{B}(\Phi_B)}\Delta^{(m)}(t, \Phi_B)\frac{\mathcal{R}(\Phi)}{\mathcal{K}(\Phi)}d\Phi.$$

Its structure naturally splits the event generation procedure into three steps: 1) generate n-body events; 2) shower the events; 3) reweight the events furnished with a shower emission.

In step 1, the n-body events need to be generated with $\widetilde{\mathcal{B}}$ as the distribution. This is equivalent with the generation of only $\mathbb{S}$-events in MC@NLO and can be achieved by setting

```
#VetoedContributionTypes
5
1 8 9 10 13
```

in the FKS_params.dat file. In addition, since a power shower will have to be performed, the normal routine must be modified to give a sufficiently large shower starting scale. The simplest way to achieve this is by setting `shower_S_scale = 1.3d4` in fks_singular.f and `shower_scale_factor = 100` in the run card. The former ensures that $\sqrt{\hat{s}_B}/2$ will not be used as the shower scale, while the latter sets the shower starting scale used to $\sim 10^4$ GeV.

In step 2, the set of events generated should be passed to PYTHIA8 for the first shower emission, generating a HepMC [25] file containing a small fraction of n-body events with no shower emission until t_{cut} and mostly $(n+1)$-body events with a shower emission.

The four-momenta of the $n+1$ particles correspond to a real phase space point Φ, which is used in step 3 to compute the multiplicative reweight factor $\mathcal{R}(\Phi)/\mathcal{K}(\Phi)$.

The procedure outlined above can be succinctly summarised as a flowchart, shown in Fig. 3a.

Note that regardless of the factorisation scheme used, the prescribed pKrkNLO procedure will be used, with $\widetilde{\mathcal{B}}$ modified accordingly.

Moreover, to obtain the Krk PDFs, the results presented in Appendix A are used to write a PYTHON script that transforms a given PDF set in the $\overline{\text{MS}}$ factorisation scheme to its Krk counterpart.

4.2 MAcNLOPS with MG5_AMC

The differential cross-section given by `MAcNLOPS` is

$$
\begin{aligned}
d\sigma_{\text{MAcNLOPS}} = &\widetilde{\mathcal{B}}(\Phi_B)\Delta(t_{\text{cut}}, \Phi_B)d\Phi_B + \\
&+ \widetilde{\mathcal{B}}(\Phi_B)\frac{\mathcal{K}}{\mathcal{B}}\Delta(t, \Phi_B)\left[1 + \frac{\mathcal{R} - \mathcal{K}}{\mathcal{K}}\Theta(\mathcal{K} - \mathcal{R})\right]d\Phi + (\mathcal{R} - \mathcal{K})\Theta(\mathcal{R} - \mathcal{K})d\Phi,
\end{aligned}
$$

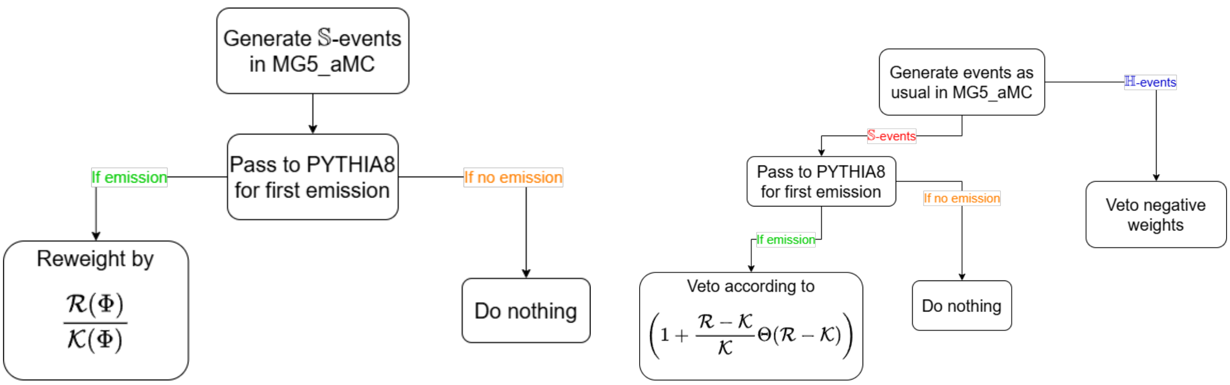
which leads to a similar split into steps as `KrkNLO`.

In step 1, events are generated with MG5_AMC as usual, giving an event sample containing both $\mathbb{S}$- and $\mathbb{H}$-events, with uniform in magnitude but possibly negative weights.

In step 2, the $\mathbb{S}$-events are passed to PYTHIA8 for the first shower emission.

Finally in step 3, the $\mathbb{H}$-events associated with a negative weight can immediately be vetoed. This is equivalent to computing $\mathcal{P}_{\text{veto}}^{(\mathbb{H})}$. For the $\mathbb{S}$-events with a shower emission, the veto probability $\mathcal{P}_{\text{veto}}^{(\mathbb{S})}$ is computed at the sampled point Φ. Then, a random number r is generated. If $\mathcal{P}_{\text{veto}}^{(\mathbb{S})}(\Phi) > r$, the event is accepted, otherwise rejected. These decisions should be recorded as `1`'s and `0`'s respectively, such that multiplication with the original weight achieves vetoing.

This procedure is summarised as a flowchart in Fig. 3b.



(a) Workflow of `pKrkNLO`

(b) Workflow of `MAcNLOPS`

4.3 Common Technicalities

Other than the procedures outlined above for `pKrkNLO` and `MAcNLOPS` respectively, some technical details common to both must be addressed as well. In particular, we discuss the settings needed for a one-emission shower, processing of the shower output, construction of Φ in the post-shower processing steps (step 3), and a deviation from MG5_AMC's `MC@NLO`-oriented standard procedure.

By default, a shower algorithm runs from the hard event scale to the hadronisation scale t_{cut}. To terminate after the first shower emission, a customised user hook must be defined, which checks the number of accepted emissions so far and vetos all subsequent attempts at generating an emission once the first is accepted. For our purpose, such user hook should be accompanied by setting `PartonLevel:FSR = off`, `PartonLevel:Remnants = off`, `Check:event = off`, `HadronLevel:all = off`, `SpaceShower:QEDshowerByQ = off` and Z^0 to be stable. With the user hook and these switches, a complete run of the shower will at most generate one ISR per event.

The output format of PYTHIA8 is by default HepMC, which contains information about the entire branching history. To extract this information, one can use the `pyhepmc` [26] PYTHON package. In particular, this gives the momenta of all $n+1$ particles involved in the one-emission process, which is the real phase space point Φ.

However, the post-shower processing need to be performed in the MG5_AMC framework, for which one cannot just use Φ as internal consistency is not guaranteed. Instead, one should use the inverse FKS mapping to obtain $\mathcal{F}_{\text{FKS}}^{-1}(\Phi) = \{\Phi'_B, \xi, y, \phi\}$ and regenerated Φ from them as is normally done in MG5_AMC. Note that this Φ'_B is generally not the same as the shower input Φ_B, but should be viewed as a fictional device allowing reconstruction of Φ in the FKS-based framework of MG5_AMC. Crucially, this means that the actual Born phase space point Φ_B must be used when computing $\mathcal{K}$.

Finally, `MG5_aMC` has a built-in mechanism that partially addresses negative $\mathbb{H}$-weights. This becomes irrelevant for `pKrkNLO` and `MAcNLOPS` and should therefore be turned off by setting `UseSudakov = .false.` in `madfks_mcatnlo.inc`.

5 Results

Following the presented implementation strategies, `pKrkNLO` and `MAcNLOPS` were implemented in MG5_AMC interfaced with PYTHIA8's default p_T-ordered shower. With `pKrkNLO`, `MAcNLOPS` and standard `MC@NLO`, event samples of 100,000 events were generated and showered. For event generation, the five-flavour scheme was employed and fixed renormalisation and factorisation scales $\mu_R = \mu_F = \sqrt{\hat{s}_B}$ are used. Since we target LHC simulation, the collision energy is set to $\sqrt{s} = 13$ TeV. For PYTHIA8, QED shower and all hadron level processes are turned off. Finally, the PDF set used throughout is CT18NLO [27] either in $\overline{\text{MS}}$ scheme or transformed to Krk scheme.

5.1 pKrkNLO

With `pKrkNLO` implemented, we obtained the results in the form of histograms for various distributions sensitive to matching schemes or valuable for practical purposes of event generators. As `pKrkNLO` provides freedom in choosing factorisation schemes, we performed two separate runs using $\overline{\text{MS}}$ and Krk schemes respectively for hard event generation[6]. Throughout this section,

[6]The shower consistently uses $\overline{\text{MS}}$ PDFs.

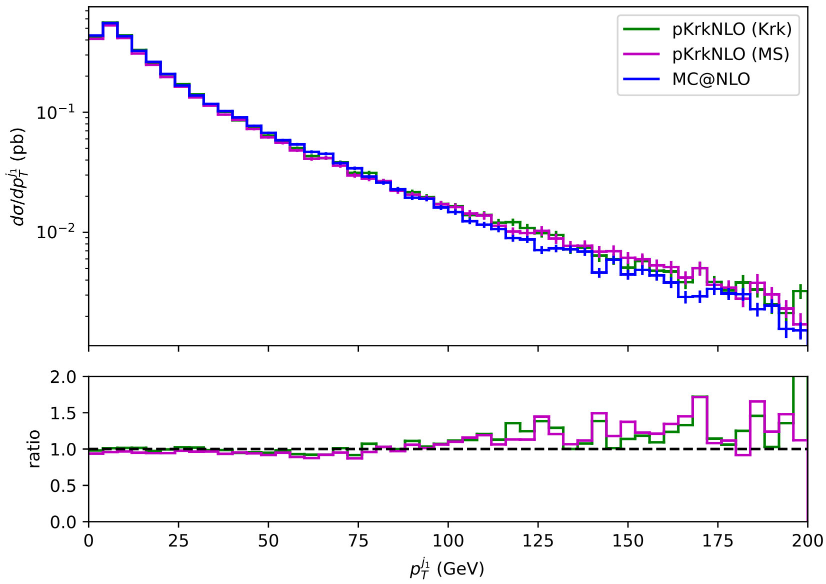


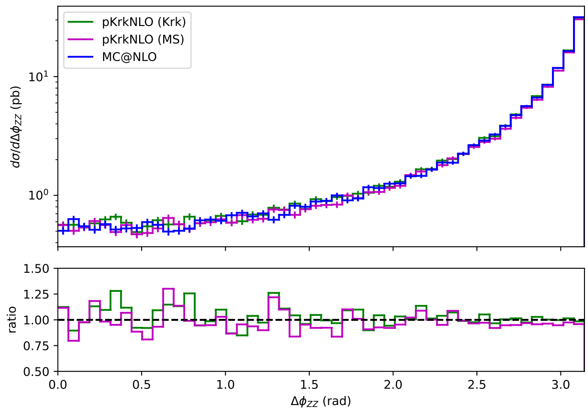
Figure 4: Histograms for the leading jet transverse momentum distribution after first emission obtained with pKrkNLO in $\overline{\text{MS}}$ and Krk PDFs, with MC@NLO as reference.

results from standard MC@NLO will be used for reference.

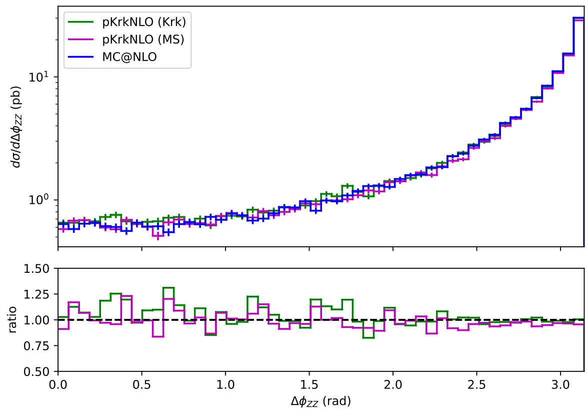
We start by presenting the histograms for the leading jet[7] transverse momentum distribution after first shower emission, shown in Fig. 4. In the same plot, we show histograms obtained from pKrkNLO with PDFs in $\overline{\text{MS}}$ (magenta) and Krk (green) schemes, along with MC@NLO (blue). For clearer comparison, the histograms are supplemented by the ratios of pKrkNLO in each scheme relative to MC@NLO. From Fig. 4, we observe that in the low-$p_T^{j_1}$ region with high statistics, the ratio is very close to one. While in the high-$p_T^{j_1}$ region, the histograms show discrepancies but continue to lie in the statistically tolerable margins. These observations imply good overall agreement.

The $p_T^{j_1}$-distributions shown in Fig. 4 is particularly interesting for us since $p_T^{j_1}$ is sensitive to soft and collinear singularities with $p_T^{j_1} \propto \xi\sqrt{1-y^2}$. As the reweight function $\mathcal{R}/\mathcal{K} \to 1$ as singularities are approached, a correct implementation would reflect this behaviour. Indeed, a visible approach to constant ratio near one is seen in Fig. 4 towards the low-$p_T^{j_1}$ region. In this region, the slight deviation from one should be attributed to the power shower, where the probability of no emission reduces with the increased range for allowed shower emissions, i.e. $\Delta^{(m)}(t_{\text{cut}}) < \Delta(t_{\text{cut}})$. The observation of the correct reweight function behaviour is an important validation of the implementation. Moreover, the deviation from one in the low-$p_T^{j_1}$ region can be observed to be smaller when Krk PDFs are used. This can only be attributed to the higher order differences introduced in the PDF transformation. Crucially, the better agreement with MC@NLO in this region does not imply that using Krk PDFs is more correct, as higher order corrections

[7]Note that the term *jet* is used loosely without a jet definition. Throughout this section, leading jet refers to the first shower emission.



(a)



(b)

Figure 5: Histograms for distribution of difference in azimuthal angle of Z^0-pair after (a) the first emission and (b) full shower for `pKrkNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference.

exist for all NLO-accurate results.

On the other hand, in the high-$p_T^{j_1}$ region, `pKrkNLO` produced histograms that generally lie above the `MC@NLO` result. This is a direct consequence of the shower then reweight procedure employed, which gave the real contribution with an additional factor $\widetilde{\mathcal{B}}/\mathcal{B} \times \Delta$. In the high-$p_T^{j_1}$ region, $\Delta \approx 1$ while $\widetilde{\mathcal{B}}/\mathcal{B} > 1$ is roughly a constant across Φ_R. Therefore, the higher bins observed in this region should be a direct consequence of $\widetilde{\mathcal{B}}/\mathcal{B} \times \Delta > 1$. Note that this is formally a higher order effect.

As for the factorisation scheme choice, it can be observed that the `pKrkNLO` histograms are effectively overlapping for the two different scheme choices. This is because the shape of the distribution is largely dependent on the shower, which consistently uses $\overline{\text{MS}}$ PDFs. Any discrepancy must then arise from the hard event generation stage, which would be higher order. Therefore, the agreement between factorisation scheme choices shows that the PDF transformation is implemented correctly.

Another distribution sensitive to the hardest (first) emission and singularities is the difference between azimuthal angles of the Z^0-pair. This is plotted in Fig. 5 for `pKrkNLO` with PDFs in both factorisation schemes compared to `MC@NLO`, with the left panel showing the distribution after first emission and right after the full shower. For matching scheme comparison, it is more instructive to examine distributions post-first emission as differences may be washed out by subsequent shower emission. On the other hand, full shower results are the ones relevant for any practical purposes and can function as a consistency check. A general examination of both Fig. 5a and 5b shows good agreement between `pKrkNLO` in either PDFs with `MC@NLO`, where the ratio is consistently observed to be compatible with one.

It is understood that $\Delta\phi_{ZZ} = \pi$ in Born events. Therefore, Fig. 5a, which is redistributed by the first emission only, has a correspondence with the $p_T^{j_1}$ distribution in Fig. 4. Namely, $\Delta\phi_{ZZ} \to \pi \iff p_T^{j_1} \to 0$. Therefore, we observed in the last bins of Fig. 5a a similar Sudakov suppression seen in the first bins of Fig. 4, where bins for `pKrkNLO` are lower than `MC@NLO` when

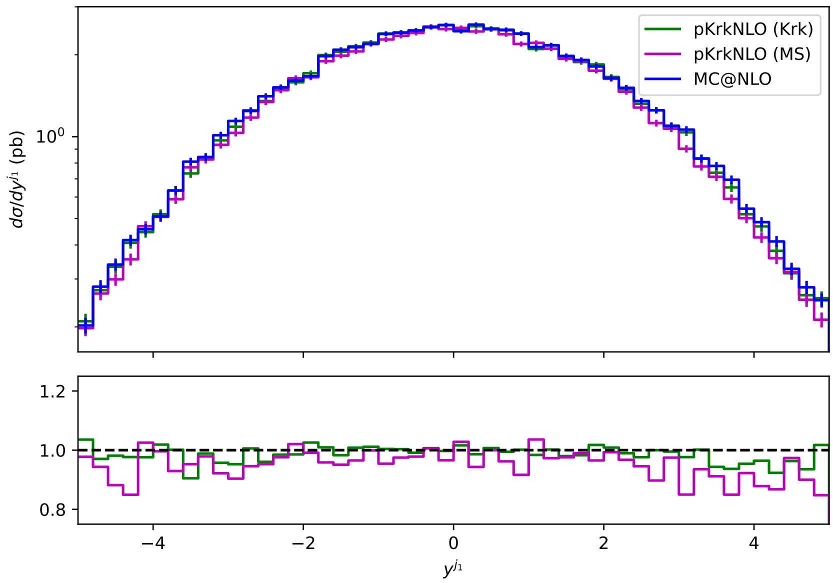


Figure 6: Histograms for the leading jet rapidity distribution after first emission obtained with pKrkNLO in $\overline{\text{MS}}$ and Krk PDFs, with MC@NLO as reference.

using $\overline{\text{MS}}$ PDFs and closer to level when using Krk PDFs. On the other hand, in remaining bins the pKrkNLO histograms fluctuate about the MC@NLO histogram.

Compared to Fig. 5a, Fig. 5b, which is redistributed by the full shower, is less sensitive to matching scheme difference. With this being said, the hardest (first) emission still has the largest impact on the distribution, as is reflected in the shape similarity between Fig. 5a and 5b. The characteristic in having Sudakov suppression in the last bins remain, while fluctuations elsewhere are moderately suppressed. Since the full shower distributions are the practically relevant ones, the good agreement with MC@NLO for $\Delta\phi_{ZZ}$ histograms after full shower further validates the pKrkNLO implementation.

We will then turn to the distribution of the leading jet rapidity y^{j_1} after first emission, plotted in Fig. 6 for pKrkNLO in $\overline{\text{MS}}$ and Krk PDFs with MC@NLO as reference. This observable is insensitive to singularities but still sensitive to matching scheme differences as its shape is determined by the first emission. Fig. 6 shows pKrkNLO and MC@NLO histograms in agreement considering statistical errors. On top of that, the symmetric shape of the y^{j_1} distribution is observed for pKrkNLO in both $\overline{\text{MS}}$ and Krk PDFs. These observations demonstrate good agreement between pKrkNLO and MC@NLO histograms, further confirming that the implementation is valid.

We can then turn to the invariant mass and rapidity separation between the Z^0-pair. These distributions are largely determined by the Born event and are therefore insensitive to matching scheme difference. However, they are of great practical value and can serve as a consistency check. Therefore, we show them in Fig. 7 for pKrkNLO in $\overline{\text{MS}}$ and Krk PDFs relative to MC@NLO. In the left (right) panels, the distributions after first emission (full shower) are shown. For all distributions plotted in Fig. 7, the pKrkNLO and MC@NLO histograms generally agree well considering statistical errors. Moreover, the shape is largely unmodified after full shower compared to first emission only, which is expected and demonstrates that the implementation has no unforeseen influence on subsequent shower.

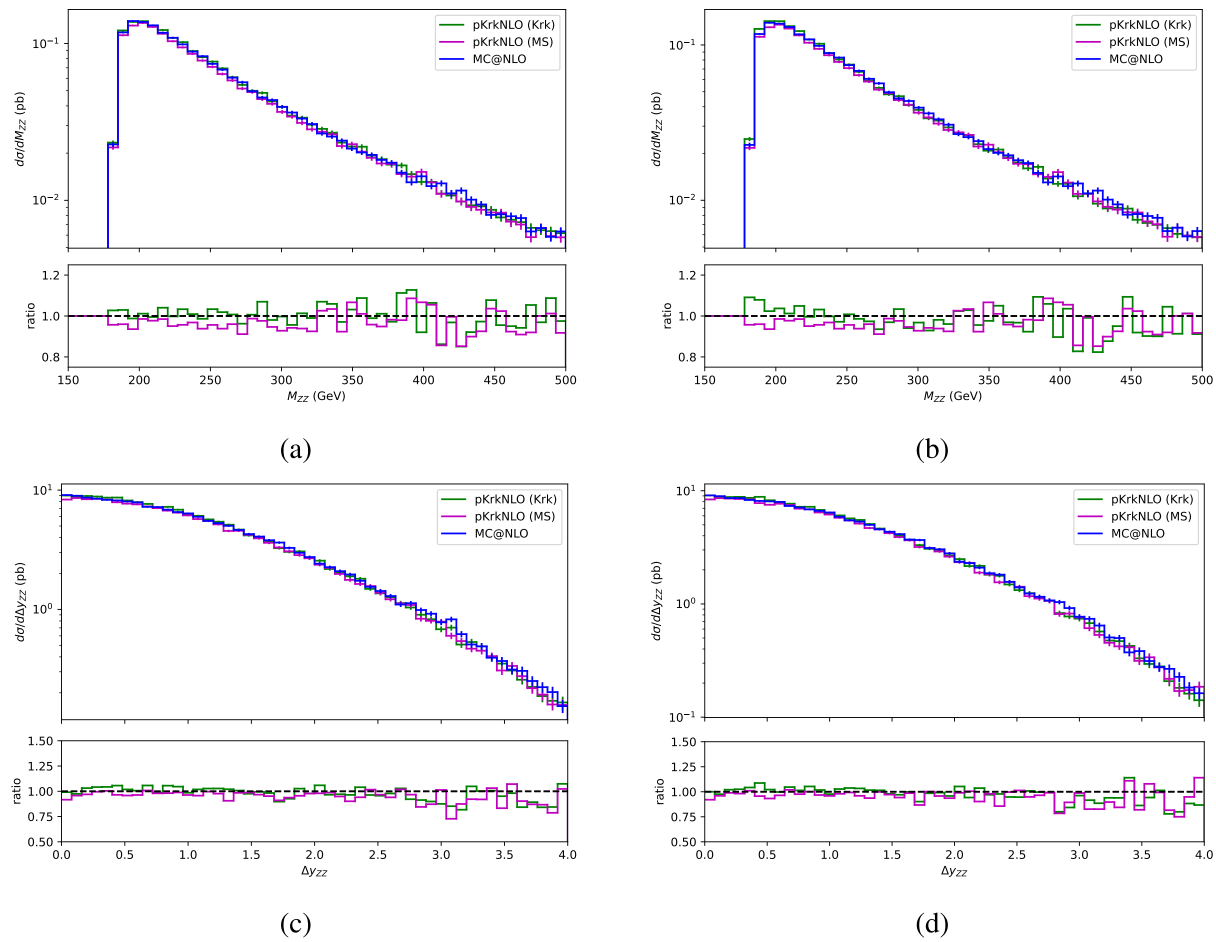
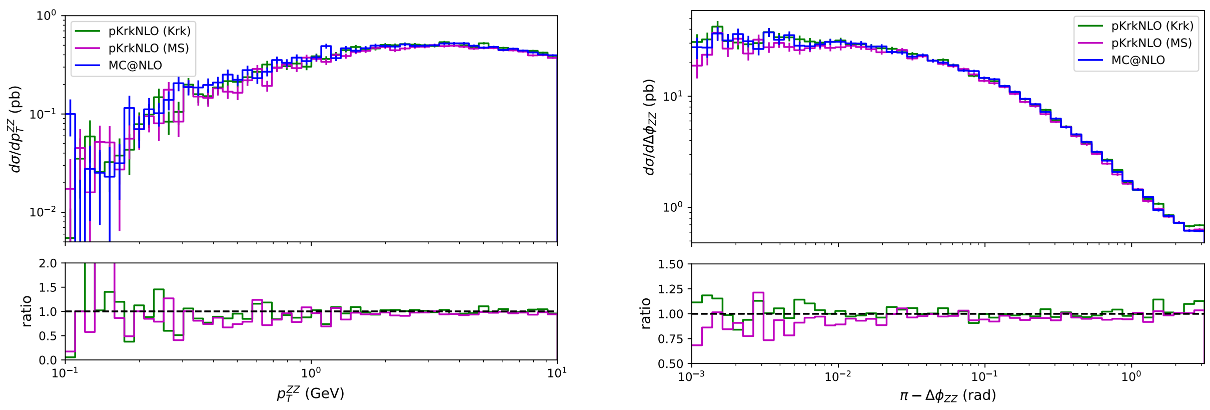


Figure 7: Histograms for (a,b) invariant mass and (c,d) rapidity separation of the Z^0-pair after first emission (left) and full shower (right) obtained with `pKrKNLO` in $\overline{\text{MS}}$ and Krk PDFs, with `MC@NLO` as reference.

Finally, we must examine the interface with hadronisation models. This concerns the shape of distributions in the regions where the transverse momentum of the Z^0-pair after full shower $p_T^{ZZ} \sim t_{\text{cut}}$, which is 0.5 GeV for default PYTHIA8 shower. In particular, if the shape does not replicate that of a showered leading-order event sample (LO+PS) up to a normalisation constant, existing hadronisation models must be retuned. It is thus desirable for the shapes to match, which `MC@NLO` accomplishes. Therefore, we can continue using `MC@NLO` as reference. In Fig. 8a, we have plotted p_T^{ZZ} on a reduced range $[0.1, 10]$ with logarithmic binning to emphasize the hadronisation region. Focusing on the low-p_T^{ZZ} region, we observe again agreement up to statistical errors. One can point to the Sudakov suppression for having a visible effect when using $\overline{\text{MS}}$ PDFs as we enter the region $p_T^{ZZ} < 1$ GeV. However, the ratio with `MC@NLO` is compatible with a constant smaller than one. Therefore, one can conclude safely that `pKrkNLO` does not



(a)

(b)

Figure 8: Histograms for (a) transverse momentum and (b) difference in azimuthal angle of Z^0-pair after full shower for pKrkNLO in $\overline{\text{MS}}$ and Krk PDFs, with MC@NLO and showered leading order as reference. Logarithmic binning is employed to emphasize hadronisation region.

require retuning the hadronisation models.

For confirmation, we have also plotted $\pi - \Delta\phi_{ZZ}$ in Fig. 8b, where the inversion of x-axis allows logarithmic binning that correctly emphasizes the hadronisation region. Despite only having the qualitative correspondence $p_T^{ZZ} \to 0 \iff (\pi - \Delta\phi_{ZZ}) \to 0$, observation of general trends can still be done. From Fig. 8b, we continue to see agreement up to statistical errors in the hadronisation region. For pKrkNLO with $\overline{\text{MS}}$ PDFs, the Sudakov suppression continues to have visible effect. For pKrkNLO with Krk PDFs, the effect is less pronounced, where the histogram appears to agree better with MC@NLO. Nevertheless, both support the previous conclusion that hadronisation models need not be retuned.

5.2 MAcNLOPS

For MAcNLOPS, we present results in a similar fashion, including distributions that are sensitive to matching scheme difference or valuable for practical usage, with MC@NLO for reference. Unlike for pKrkNLO, there is no incentive to use the Krk factorisation scheme. Therefore, all results are obtained with the standard $\overline{\text{MS}}$ PDFs. Note also that MAcNLOPS does not require power shower.

We begin with the leading jet transverse momentum distribution after first emission, plotted in Fig. 9. The MAcNLOPS and MC@NLO histograms are plotted in red and blue respectively. Note that MAcNLOPS features reduced fluctuations in the high-$p_T^{j_1}$ region. This is a direct consequence of the fact that the sampling technique in the high-$p_T^{j_1}$ region is the same for MC@NLO and MAcNLOPS[8]. The fluctuations in Fig. 9 are thus purely statistical, as opposed to for pKrkNLO, where both the power shower and reweighting contribute to large fluctuations. From Fig. 9, we observe that the histograms are effectively overlapping in the high-statistics region and remains in agreement

[8]The shower operates identically, dropping out due to the dead zone setting discussed in Sect. 3.2, and negative $\mathbb{H}$-weights typically do not appear in the high-$p_T^{j_1}$ region.

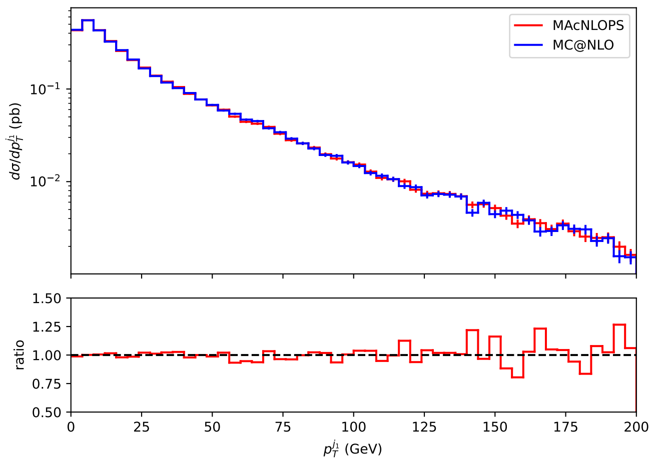


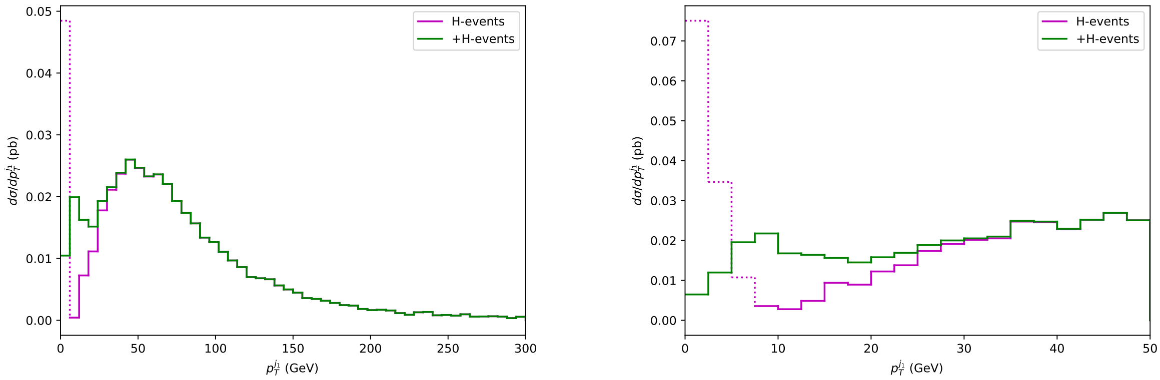
Figure 9: Histograms for the leading jet transverse momentum distribution after first emission obtained with `MAcNLOPS`, with `MC@NLO` as reference.

considering statistical errors in low-statistics region. This gives ratio consistently compatible with one, demonstrating great agreement between `MAcNLOPS` and `MC@NLO`. One should note that the additional Sudakov suppression upon `MC@NLO` is arguably non-existent, since power shower is no longer employed.

To better understand the underlying mechanism of `MAcNLOPS`, it is instructive to examine the separate effects of negative-$\mathbb{H}$ removal and `MAcNLOPS`. To this end, we must first explore the distribution of negative $\mathbb{H}$-weights. This is plotted in Fig. 10, featuring only $\mathbb{H}$-events (magenta) or positive $\mathbb{H}$-events (green) from the `MC@NLO` event sample used for `MAcNLOPS`. Note that such an `MC@NLO` event sample is generated with `UseSudakov = .false.` and is not the standard `MC@NLO` used for reference throughout the Results section. The negative bins are inverted and labelled with dotted lines. From Fig. 10a, one can see that the region where $p_T^{j_1} \gtrsim 50$ GeV is mostly free of negative-$\mathbb{H}$ contribution. Therefore, Fig. 10b is zoomed in on the range $p_T^{j_1} \in [0, 50]$ GeV. From Fig. 10b, we observe negative $\mathbb{H}$-contribution to start being of comparable size with positive $\mathbb{H}$-contribution at $p_T^{j_1} \approx 10$ GeV and dominates as we move to $p_T^{j_1} \to 0$.

Since the `MAcNLOPS` vetoing is active only in negative $\mathbb{H}$-regions, we should restrict the $p_T^{j_1}$ range for studies of its effect. In Fig. 11, this range is chosen to be $p_T^{j_1} \in [0, 25]$ GeV. Fig. 11 shows the effects of negative $\mathbb{H}$-removal and `MAcNLOPS` vetoing separately, while providing the final `MAcNLOPS` and input `MC@NLO` as references. The ratios are consistently relative to the input `MC@NLO`, which we emphasize again is not the standard `MC@NLO`. From Fig. 11, we observe in all bins except the first few complementary effects from removing negative $\mathbb{H}$-events and performing `MAcNLOPS` vetoing. The range that these bins correspond to is $p_T^{j_1} \in [1.5, 25]$ GeV, where negative $\mathbb{H}$-events are sizeable but not dominating. In the first few bins ($p_T^{j_1} \in [0, 1.5]$ GeV), where negative $\mathbb{H}$-events do dominate, we observe the removal of negative $\mathbb{H}$-events to overwhelm `MAcNLOPS` vetoing, especially in the first bin.

We can investigate this discrepancy by direct inspection of the negative-$\mathbb{H}$ contribution removed and the veto-related contribution in `MAcNLOPS`. These can be found implicitly in the



(a) (b)

Figure 10: Histograms of only $\mathbb{H}$-event and positive $\mathbb{H}$-event contribution to the leading jet transverse momentum distribution for two different $p_T^{j_1}$ ranges. Negative bins are reflected and plotted in dotted lines.

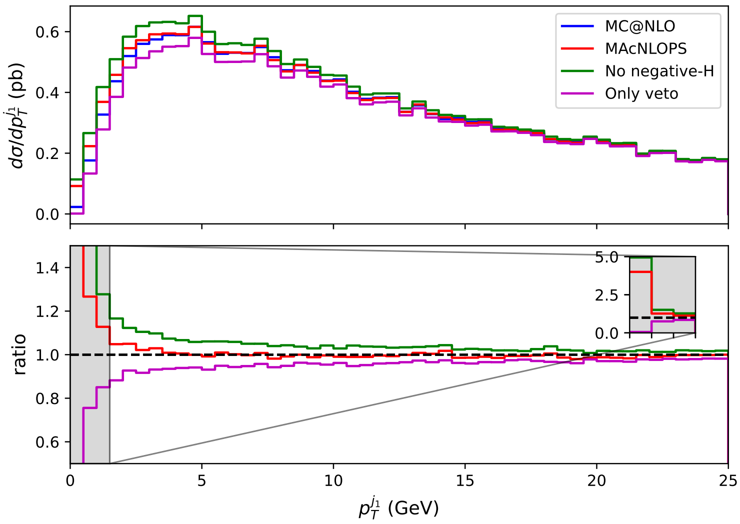
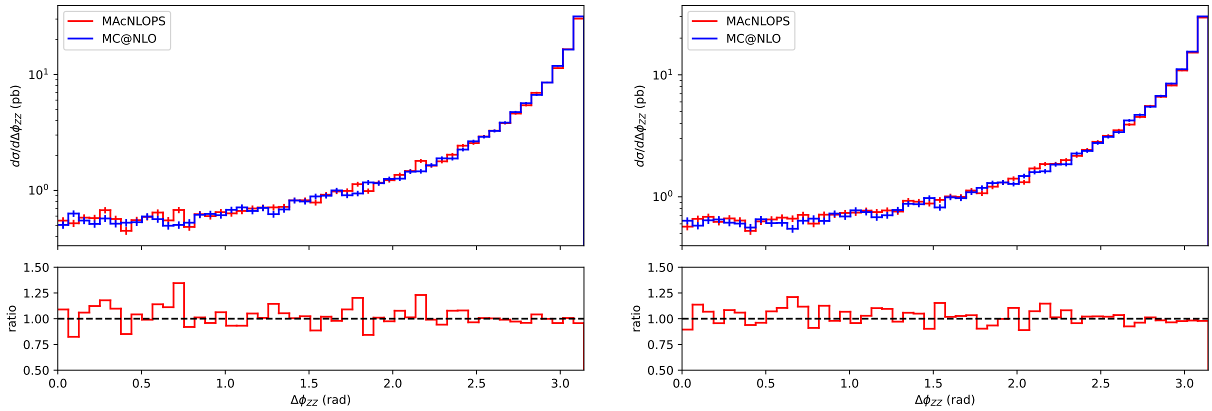


Figure 11: Histograms showing the separate effects of removing negative $\mathbb{H}$-events and `MAcNLOPS` vetoing, with `MAcNLOPS` and its input `MC@NLO` for reference.

proof of NLO matching condition shown in Sect. 3.4. The removed $\mathbb{H}$-contribution is $(\mathcal{R} - \mathcal{K})\Theta(\mathcal{K} - \mathcal{R})$ while the veto-related contribution is $(\widetilde{\mathcal{B}}/\mathcal{B}) \times \Delta \times (\mathcal{R} - \mathcal{K})\Theta(\mathcal{K} - \mathcal{R})$. Towards $p_T^{j_1} \approx 0$, the Sudakov damping is non-negligible as the probability of no emission is low, giving $(\widetilde{\mathcal{B}}/\mathcal{B}) \times \Delta < 1$. Therefore, the removed negative $\mathbb{H}$-contribution will not be fully restored by `MAcNLOPS` due to the higher order Sudakov damping. On top of this, the first bin, corresponding to range $p_T^{j_1} \in [0, 0.5]$ GeV, lies close to the regions inaccessible by the shower due to cut-off. This further enhances the undersampling problem, which is observed as the sudden spike in effect of negative-$\mathbb{H}$ removal in the first bin. Moreover, in practice we cannot overlook negative $\mathbb{S}$-weights, which makes vetoing increase the bin heights instead of reducing them. These factors combined led to the observation in Fig. 11, where negative-$\mathbb{H}$ removal overwhelmed MAcNLOPS vetoing.



(a) (b)

Figure 12: Histograms for distribution of difference in azimuthal angle of Z^0-pair after (a) the first emission and (b) full shower for MAcNLOPS, with MC@NLO as reference.

We can now turn to the other distribution sensitive to singularities, which is the $\Delta\phi_{ZZ}$ distribution plotted in Fig. 12b for after first emission (left) and full shower (right). As usual, we show MAcNLOPS with standard MC@NLO as reference. In both Fig. 12a and 12b, good agreement was observed between the histograms considering statistical errors. Moreover, Fig. 12a confirms the previously observed behaviour in Fig. 9, showing no consistent additional Sudakov suppression pattern in the $\Delta\phi_{ZZ} \approx \pi$ bins. The similarity between Fig. 12a and 12b is again expected, providing support for validation of the implementation.

One more distribution sensitive to matching schemes is the leading jet rapidity, plotted in Fig. 13 for MAcNLOPS with MC@NLO as reference. We see in Fig. 13 again histograms in agreement up to statistical errors. Across the range plotted, the ratio is generally fluctuating, appearing more stable in region with higher statistics and less in lower. This pattern is typical when there are no systematic differences, hence validates the implementation.

Similar to the preceding section, we will now move on to inclusive distributions in M_{ZZ} and Δy_{ZZ}, both of which have shapes largely dependent on the Born event. The histograms for after first shower emission (left) and after full shower (right) obtained with MAcNLOPS compared to MC@NLO are shown in Fig. 14. All histograms are in good agreement considering statistical errors. Moreover, the low sensitivity to emissions are correctly represent by the similarity between distributions after first emission and full shower. These observations validate the consistency of the implementation. Since these distributions will be of practical value, one can safely say that MAcNLOPS is indistinguishable from MC@NLO for practical purposes.

Finally, we must still check compatibility with hadronisation models. To this end, histograms focusing on the hadronisation region where $p_T^{ZZ} \lesssim t_{\text{cut}}$ are plotted in Fig. 15 with logarithmic

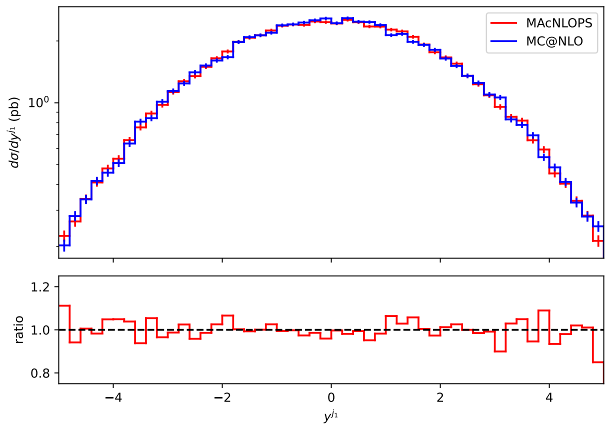


Figure 13: Histograms for the leading jet rapidity distribution after first emission obtained with `MAcNLOPS`, with `MC@NLO` as reference.

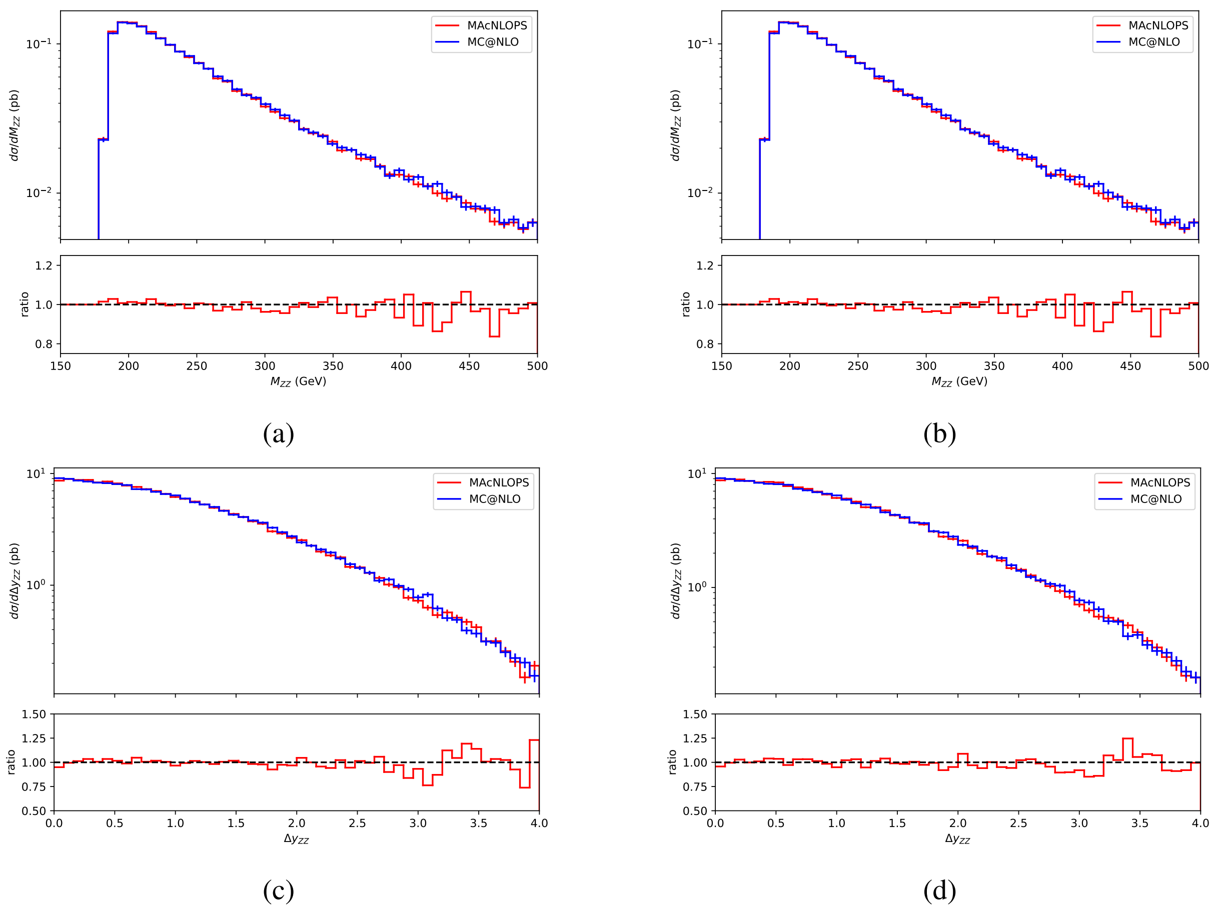


Figure 14: Histograms for (a,b) invariant mass and (c,d) rapidity separation of the Z^0-pair after first emission (left) and full shower (right) obtained with `MAcNLOPS`, with `MC@NLO` as reference.

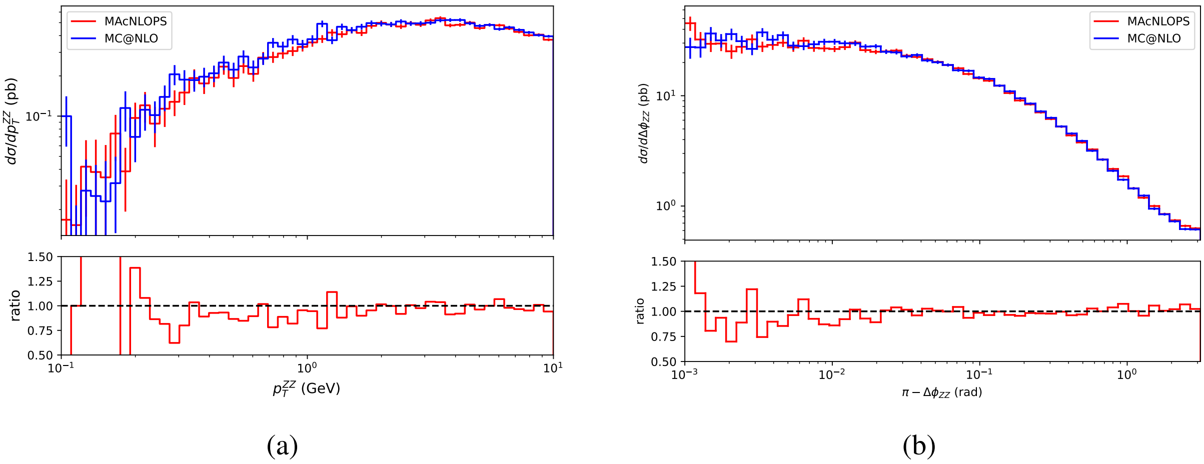


Figure 15: Histograms for (a) transverse momentum and (b) difference in azimuthal angle of Z^0-pair after full shower for `MAcNLOPS`, with `MC@NLO` and showered leading order as reference. Logarithmic binning is employed to emphasize hadronisation region.

binning for leading jet transverse momentum and azimuthal angle difference between Z^0-pair after full shower. For p_T^{ZZ} in Fig. 15a, the range is adjusted to $[0.1, 10]$ since high-p_T^{ZZ} is irrelevant for hadronisation model. While for $\Delta\phi_{ZZ}$, $(\pi - \Delta\phi_{ZZ})$ was plotted instead. We should note also the reduction of y-range from $[0, 2]$ to $[0.5, 1.5]$ compared to the `pKrkNLO` counterpart in Fig. 8. Looking at p_T^{ZZ} first, a generally good agreement is observed. This is supported by the $\Delta\phi_{ZZ}$ histogram as well. Therefore, one should make the conclusion that `MAcNLOPS` follows the `MC@NLO` shape and would thus not require retuning the hadronisation model.

6 Conclusion

In this thesis, we presented the implementations of `KrkNLO` and `MAcNLOPS` using MG5_aMC for $pp \to Z^0 Z^0$. This allowed an investigation of their performances relative to `MC@NLO`, the standard NLO matching scheme. In addition, the proposed implementation (`pKrkNLO`) no longer restricts `KrkNLO` to using only the Krk factorisation scheme, whose effect is also studied.

The results are presented as a set of histograms for jet and inclusive distributions, including ones sensitive to performances of matching schemes and others that are of practical importance. Generally, for both `KrkNLO` (in $\overline{\text{MS}}$ and Krk factorisation schemes) and `MAcNLOPS`, good agreement with `MC@NLO` was observed, up to higher order discrepancies, which validates the implementations. Moreover, the computational cost required to perform the post-shower processing step (reweight for `KrkNLO`, veto for `MAcNLOPS`) are confirmed to be negligible as phase space sampling was not involved. Therefore, the advantage of both matching schemes in generating a positive-definite event sample can be said to come with no extra cost.

For `KrkNLO`, due to the need to use a power shower and multiplicative reweighting, the fluctuations are generally larger than `MAcNLOPS`, whose discrepancies with `MC@NLO` are largely statistical. This issue can be overcome by generating a larger event sample and performing extra unweighting, but would partially offset the gain from having a positive-definite event sample. In addition, KrkNLO introduces inherent higher order discrepancies, in the form of additional Sudakov suppression in the near-singular regions and shower oversampling in the high-p_T tail. This, however, should not necessarily be viewed as a disadvantage since the discrepancies are beyond NLO accuracy.

KrkNLO is also observed to be mostly insensitive to factorisation scheme choice. This is largely due to the use of $\overline{\text{MS}}$ PDFs in the shower regardless of factorisation scheme choice at hard event generation, which was the conventional approach. Formally, there is a mismatch between parton shower and hard event generation when Krk factorisation scheme is chosen. However, such a mismatch would introduce only higher order corrections[9] and thus does not represent a problem for NLO accuracy.

In the hadronisation region, Sudakov suppressions were observed for KrkNLO. The suppression was greater when using $\overline{\text{MS}}$ PDFs and smaller when using Krk PDFs, which is speculated to be due to higher order discrepancies introduced when employing Krk PDFs compensating for the suppression. If this is the case, one cannot assume that such compensation will happen generally. Therefore, it is most logical to say that KrkNLO, regardless of factorisation scheme, will be subject to a Sudakov suppression due to using a power shower. However, this does not lead to the need of retuning hadronisation models as the suppression due to power shower makes the distribution follow the LO+PS results instead. For MAcNLOPS, the behaviour in hadronisation region is compatible with MC@NLO results, also indicating that no retuning is needed.

In terms of applicability, we expect the presented implementation to be directly transferable to all colour-singlet processes with $q\bar{q}$ initial state at leading order, whereas for more complex processes, a careful revision is required. Such a revision would necessarily involve further clarification of phase space mappings, especially when FSRs are included. For MAcNLOPS, it may additionally be desirable for further implementations to use MC@NLO with some negative $\mathbb{S}$-weight reduction method (e.g. [20–23]) for hard event generation to reduce interference with the vetoing procedure.

By performing the implementations proposed, we were able to observe positive results demonstrating that the novel matching schemes KrkNLO and MAcNLOPS can enter the conversation as alternatives of established matching schemes. As both matching schemes remove the undesirable negative weights at negligible cost, one can imagine scenarios where they would be preferred over MC@NLO. This is especially the case for MAcNLOPS, which is capable of generating unweighted event samples on-the-fly. With a full implementation of KrkNLO already in progress [9], we believe there is an argument to be made for the same being done for MAcNLOPS.

Finally, with the development of event generators proceeding towards higher precision, an immediate step to consider is matching to NLL shower, i.e. NLO+NLL. Such an extension should be straightforward for both KrkNLO and MAcNLOPS [8]. From this perspective, they should be examined in parallel with the recently-proposed class of Exponentiated Subtraction for Matching Events (ESME) schemes [28], which are constructed for NLO+NLL and also guarantee positive-definite event generation. This could be the subject of further studies.

[9] In fact, as the PDF evolution equations also agree only up to higher order effects for different factorisation schemes, using Krk PDFs will also introduce a higher order discrepancy.

Acknowledgments

I'm extremely grateful to my supervisor Rikkert for introducing me to this exciting topic, trusting my ability and providing guidance throughout. He was always patient and encouraging, even when the progress was stagnating and the direction was somewhat lost. His help was crucial for the completion of this thesis and will continue to be invaluable in my future endeavours. Thank you!

I would also like to thank my family, from whom I got constant emotional support, which allowed me to always face next day's challenge without carrying baggage from the previous day. Much of my strength comes from their belief in me.

Finally, I would like to thank my friends, whose presence is a great reminder that a task at hand, however important, can never be all that there is to life and should not stop you from enjoying it.

A Transformation to Krk PDFs

We present here the derivation leading to the two equations used to transform PDFs from the standard $\overline{\text{MS}}$ factorisation scheme to Krk scheme. This section is built on Sect. 2.3 and Appendix A in [9].

To NLO in α_S, the PDFs in a new factorisation scheme (in our case Krk) are related to the $\overline{\text{MS}}$ PDFs by

$$f_a^{\text{Krk}}(x,\mu_F) = f_a^{\overline{\text{MS}}}(x,\mu_F) + \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\sum_b K_{ab}^{\overline{\text{MS}}\to\text{Krk}}(z) f_b^{\overline{\text{MS}}}\left(\frac{x}{z},\mu_F\right), \tag{A.1}$$

where $K_{ab}^{\overline{\text{MS}}\to\text{Krk}}(z)$ is a *transformation kernel.*

By defining such a new factorisation scheme, we wish to eliminate the collinear remnant terms $\mathcal{C}_\pm$. To this end, we must write down its specific form, which is most compact in the Catani-Seymour formalism [29]

$$\mathcal{C}(x,\mu_F) = P_{ab}(x,\mu_F) + \underbrace{\bar{K}_{ab}(x) + \widetilde{K}_{ab}(x)}_{\text{K}_{ab}^{\overline{\text{MS}}}(x)} - K_{ba}^{\text{FS}}(x),$$

where P_{ab} contains the factorisation scale dependence while $\bar{K}_{ab}$ and $\widetilde{K}_{ab}$ relate to the factorisation of initial-state collinear singularities into PDFs. The final term K_{ab}^{FS} is factorisation scheme dependent, given by the finite part also factorised away. Therefore, the $\overline{\text{MS}}$ scheme can be thought of as defined by $K_{ab}^{\overline{\text{MS}}} = 0$. In the followed derivations, we redefine this term to be $K_{ab}^{\overline{\text{MS}}\to\text{Krk}}$ to be more explicit about the function of this term.

Since $P(x,\mu_F = \hat{s}) = 0$, it is reasonable to define the transformation kernel in a general form of

$$K_{ab}^{\overline{\text{MS}}\to\text{Krk}}(x) = \text{K}_{ba}^{\overline{\text{MS}}}(x) - \Delta^{\text{Krk}}\delta(1-x), \tag{A.2}$$

where the δ-function term helps satisfy the imposed momentum sum rule and, upon integration, can be combined with the $\widetilde{\mathcal{B}}$ contribution.

Therefore, if we set the factorisation scale $\mu_F = \hat{s}$ and use the Krk factorisation scheme, we get

$$P_{ab}(x, \mu_F = \sqrt{\hat{s}}) + \mathrm{K}^{\mathrm{Krk}}_{ab}(x) = \Delta^{\mathrm{Krk}}\delta(1-x), \tag{A.3}$$

which is clearly a great simplification, allowing the collinear remnant term to be eliminated.

The distinct cases we need to consider are just $ab = \{qq, qg, gg\}$, which are

$$K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{qq}(x) = \mathrm{K}^{\overline{\mathrm{MS}}}_{qq}(x) - \frac{1}{2}\Delta^{\mathrm{Krk}}_0\delta(1-x) \tag{A.4}$$

$$K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{qg}(x) = \mathrm{K}^{\overline{\mathrm{MS}}}_{gq}(x) \tag{A.5}$$

$$K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{gq}(x) = \mathrm{K}^{\overline{\mathrm{MS}}}_{qg}(x) \tag{A.6}$$

$$K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{gg}(x) = \mathrm{K}^{\overline{\mathrm{MS}}}_{gg}(x) - \Delta^{\mathrm{Krk}}_1\delta(1-x), \tag{A.7}$$

where the extra $1/2$ in K_{qq} is added for agreement with [9]. Note that including the delta function in two kernels will be sufficient to satisfy momentum sum rules, so qq and gg has been chosen.

The remaining task is to simply plug in explicit expressions (available in Appendix C of [29]) and determine the free constants Δ^{Krk} from the imposed sum rule. We omit the full derivation and just present the final simplifications leading up to the result, which are

$$\begin{aligned}
K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{qq}(x) &= C_F\Bigg(4\left[\frac{\log(1-x)}{1-x}\right]_+ - 2\left[\frac{\log(x)}{1-x}\right]_+ - (1+x)\log\frac{(1-x)^2}{x} + \\
&\qquad + 1 - x - \frac{17}{4}\delta(1-x)\Bigg) \\
&= C_F\Bigg(2\left[\frac{1}{1-x}\log\frac{(1-x)^2}{x}\right]_+ - \frac{1-x^2}{1-x}\log\frac{(1-x)^2}{x} + 1 - x - \frac{17}{4}\delta(1-x)\Bigg) \\
&= C_F\Bigg(\left[\frac{1+x^2}{1-x}\log\frac{(1-x)^2}{x} + 1 - x\right]_+ - \frac{17}{4}\delta(1-x) - \\
&\qquad - \delta(1-x)\int_0^1 dz\left\{(1+z)\log\frac{(1-z)^2}{z} - 1 + z\right\}\Bigg) \\
&= C_F\Bigg(\left[\frac{1+x^2}{1-x}\log\frac{(1-x)^2}{x} + 1 - x\right]_+ - \frac{3}{2}\delta(1-x)\Bigg) \\
K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{qg}(x) &= T_R\Bigg([x^2 + (1-x)^2]\log\frac{(1-x)^2}{x} + 2x(1-x)\Bigg) \\
K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{gq}(x) &= C_F\Bigg(\frac{1+(1-x)^2}{x}\log\frac{(1-x)^2}{x} + x\Bigg) \\
K^{\overline{\mathrm{MS}}\to\mathrm{Krk}}_{gg}(x) &= C_A\Bigg(2\left[\frac{1}{1-x}\log\frac{(1-x)^2}{x}\right]_+ + 2\left(\frac{1-x}{x} - 1 + x(1-x)\right)\log\frac{(1-x)^2}{x} - \\
&\qquad - \delta(1-x)\left(\frac{341}{72} - \frac{59}{36}\frac{n_f T_R}{C_A}\right)\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= C_A\Bigg(4\left[\frac{\log(1-x)}{1-x}\right]_+ - 2\frac{\log x}{1-x} + 2\left(\frac{1}{x} - 2 + x(1-x)\right)\log\frac{(1-x)^2}{x} - \\
&\qquad - \delta(1-x)\left(\frac{341}{72} - \frac{59}{36}\frac{n_f T_R}{C_A} - 2\int_0^1 dx \frac{\log x}{1-x}\right)\Bigg) \\
&= C_A\Bigg(4\left[\frac{\log(1-x)}{1-x}\right]_+ - 2\frac{\log x}{1-x} + 2\left(\frac{1}{x} - 2 + x(1-x)\right)\log\frac{(1-x)^2}{x} - \\
&\qquad - \delta(1-x)\left(\frac{341}{72} - \frac{59}{36}\frac{n_f T_R}{C_A} + \frac{\pi^2}{3}\right)\Bigg),
\end{aligned}
$$

where we have used

$$
\begin{aligned}
\Delta_0^{\text{Krk}} &= C_F\left(2\pi^2 - \frac{3}{2}\right) \\
\Delta_1^{\text{Krk}} &= C_A\left(\frac{2\pi^2}{3} - \frac{59}{72} + \frac{5}{36}\frac{T_R n_f}{C_A}\right).
\end{aligned}
$$

To complete the set we note that

$$
\begin{aligned}
K_{\bar{q}\bar{q}}^{\overline{\text{MS}}\to\text{Krk}} &= K_{qq}^{\overline{\text{MS}}\to\text{Krk}} \quad, K_{\bar{q}q}^{\overline{\text{MS}}\to\text{Krk}} = K_{q\bar{q}}^{\overline{\text{MS}}\to\text{Krk}} = 0, \\
K_{g\bar{q}}^{\overline{\text{MS}}\to\text{Krk}} &= K_{gq}^{\overline{\text{MS}}\to\text{Krk}} \quad, K_{\bar{q}g}^{\overline{\text{MS}}\to\text{Krk}} = K_{qg}^{\overline{\text{MS}}\to\text{Krk}}.
\end{aligned}
$$

These results replicate the ones obtained in [11].

The PDF transformations are

$$
\begin{aligned}
f_q^{\text{Krk}}(x,\mu_F) =& f_q^{\overline{\text{MS}}}(x,\mu_F) - \frac{\alpha_S}{2\pi}\frac{3}{2}C_F f_q^{\overline{\text{MS}}}(x,\mu_F) \; + \\
&+ \frac{\alpha_S}{2\pi}C_F\int_x^1 \frac{dz}{z}\left[\frac{1+z^2}{1-z}\log\frac{(1-z)^2}{z} + 1 - x\right]_+ f_q^{\overline{\text{MS}}}\left(\frac{x}{z},\mu_F\right) + \\
&+ \frac{\alpha_S}{2\pi}T_R\int_x^1 \frac{dz}{z}\left([z^2+(1-z)^2]\log\frac{(1-z)^2}{z} + 2z(1-z)\right) f_g^{\overline{\text{MS}}}\left(\frac{x}{z},\mu_F\right)
\end{aligned} \tag{A.8}
$$

$$
\begin{aligned}
f_g^{\text{Krk}}(x,\mu_F) =& f_g^{\overline{\text{MS}}}(x,\mu_F) - \frac{\alpha_S}{2\pi}C_A\left(\frac{341}{72} - \frac{59}{36}\frac{n_f T_R}{C_A} + \frac{\pi^2}{3}\right) f_g^{\overline{\text{MS}}}(x,\mu_F) \; + \\
&+ \frac{\alpha_S}{2\pi}C_A\int_x^1 \frac{dz}{z}\Bigg(4\left[\frac{\log(1-z)}{1-z}\right]_+ - 2\frac{\log z}{1-z} \\
&\qquad + 2\left(\frac{1}{z} - 2 + z(1-z)\right)\log\frac{(1-z)^2}{z}\Bigg) f_g^{\overline{\text{MS}}}\left(\frac{x}{z},\mu_F\right) \\
&+ \frac{\alpha_S}{2\pi}C_F\sum_{q,\bar{q}}\int_x^1 \frac{dz}{z}\left(\frac{1+(1-z)^2}{z}\log\frac{(1-z)^2}{z} + z\right) f_q^{\overline{\text{MS}}}\left(\frac{x}{z},\mu_F\right).
\end{aligned} \tag{A.9}
$$

Note that we can use the plus-prescription despite the integration limits being $[x,1]$ as $z < x \implies x/z > 1$, where the PDFs vanish.

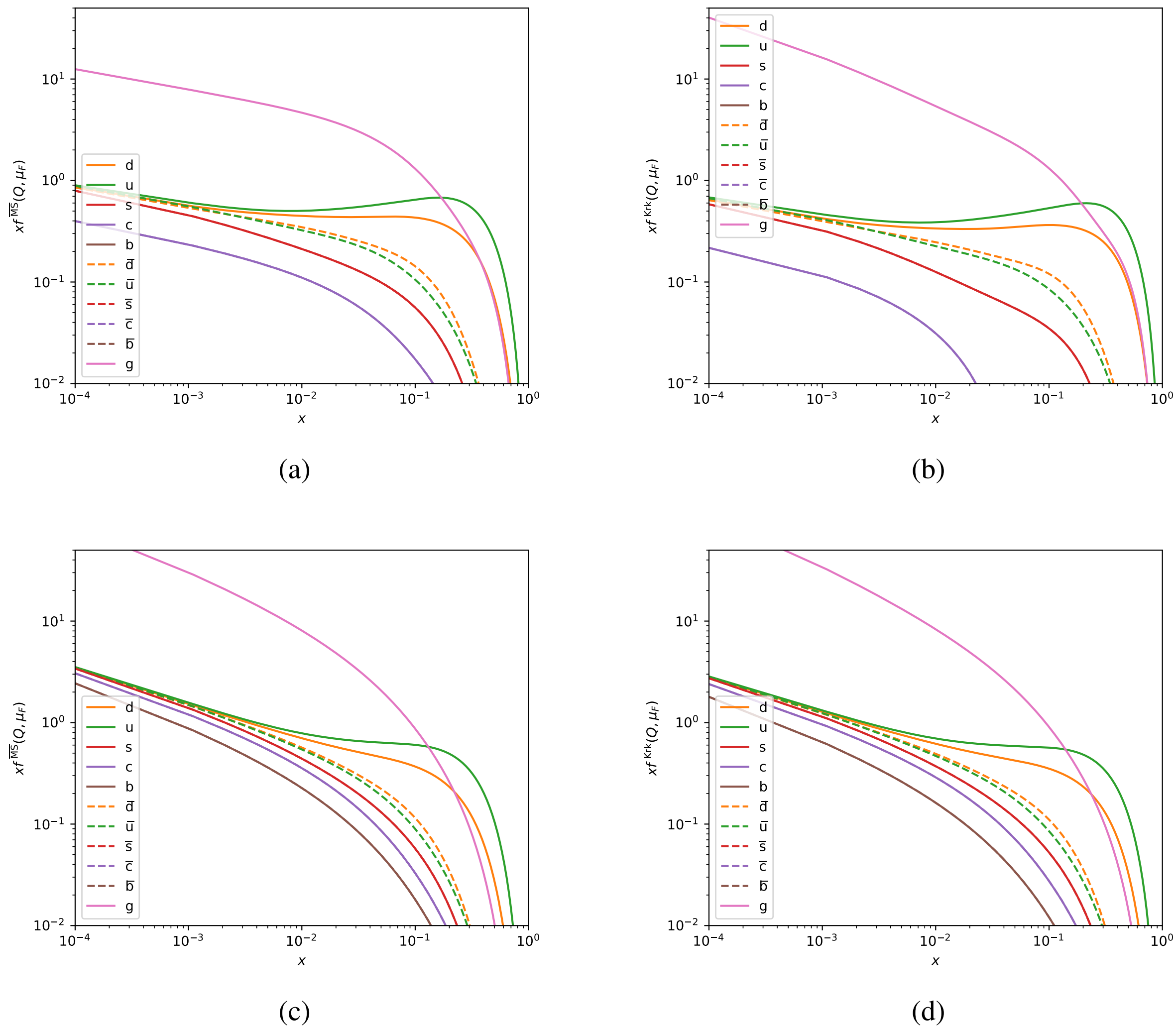


Figure 16: Plots of $\overline{\text{MS}}$ and Krk PDFs at scales 3 GeV (top) and 100 GeV (bottom).

To display the differences between the transformed PDFs in the Krk scheme and original PDFs in the $\overline{\text{MS}}$ scheme, we show in Fig. 16 a plot of Bjorken x against $x \cdot f(x, \mu_F)$ at two scales 3 and 100 GeV for the full set of PDFs in both schemes.

It is evident from Fig. 16 that the difference lies mainly in the enhanced gluon PDF in the Krk scheme, compensated by suppressed quark PDFs. This difference is, however, less pronounced as the scale increases from 3 to 100 GeV. The difference also appears to be more significant in the low-x region, while remaining visible but becomes much smaller at the moderate to high-x region. A more in-depth study into the differences induced by factorisation scheme choice can be found in Ref. [30].

B Event Samples with Different Weight Structures

Consider an event sample of size n that is positive-definite with a uniform weight associated to each event. Then we can without loss of generality assume the weight to be one, giving

$$\sum_i^n w_i = n \text{ and } \sum_i^n w_i^2 = n.$$

The quantity we are interested in when it comes to precision is the relative error r, defined as

$$r^2 \equiv \frac{\sum_i^n w_i^2}{\left(\sum_i^n w_i\right)^2}.$$

Therefore, for a positive-definite and unweighted event sample of size n, the relative error is given by

$$r^2 = \frac{1}{n} \tag{B.1}$$

Consider now another event sample that is unweighted but contains a fraction f of negative events. Let this event sample be of size N, then we have

$$\sum_i^N w_i = (1-f)N - fN = (1-2f)N \text{ and } \sum_i^N w_i^2 = N,$$

since $(1-f)N$ events are positively weighted and fN events are negatively weighted.

The relative error, which we choose to denote by r_{neg}, is then given by

$$r_{\text{neg}}^2 = \frac{1}{(1-2f)^2 N}. \tag{B.2}$$

Therefore, if we wish to have this event sample be at the same level of precision as the positive definite and unweighted event sample, it is required that

$$N = \frac{n}{(1-2f)^2} \tag{B.3}$$

In addition, it is also possible to have a weighted but positive definite event sample. For such an event sample, its convenient to start with the variance σ^2, given by

$$\sigma^2 = \sum_i^M w_i^2 - \frac{\left(\sum_i^M w_i\right)^2}{M} \geq 0,$$

where the equality holds by definition and we denote the sample size by M.

The relative error, which we choose to denote by r_{weighted}, can now be rewritten as

$$r_{\text{weighted}}^2 = \frac{\sum_i^M w_i^2}{\left(\sum_i^M w_i\right)^2} = \frac{\sigma^2 + \left(\sum_i^M w_i\right)^2/M}{\left(\sum_i^M w_i\right)^2} = \frac{1}{M} + \frac{\sigma^2}{\left(\sum_i^M w_i\right)^2} \geq \frac{1}{M}, \tag{B.4}$$

where equality only happens when the events are unweighted.

If the sample size is the same with a positive-definite and unweighted event sample, $M = n \implies r^2_{\text{weighted}} > r^2$, demonstrating that having weighted events lead to greater imprecision. Therefore, a larger event sample is also required to have the same precision.

This can be interpreted as $1/r^2$ being an effective sample size, where having negative or non-uniform weight both lead to a reduction of effective sample size. Such interpretation led to the definition of the *Kish effective sample size* [31].

References

[1] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [2203.11601].

[2] J. Bellm et al., *The Physics of Herwig 7*, 2512.16645.

[3] SHERPA collaboration, *Event generation with Sherpa 3*, *JHEP* **12** (2024) 156 [2410.22148].

[4] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [hep-ph/0204244].

[5] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].

[6] S. Jadach, W. Płaczek, S. Sapeta, A. Siódmok and M. Skrzypek, *Matching NLO QCD with parton shower in Monte Carlo scheme — the KrkNLO method*, *JHEP* **10** (2015) 052 [1503.06849].

[7] P. Nason and G.P. Salam, *Multiplicative-accumulative matching of NLO calculations with parton showers*, *JHEP* **01** (2022) 067 [2111.03553].

[8] K. Hamilton, A. Karlberg, G.P. Salam, L. Scyboz and R. Verheyen, *Matching and event-shape NNDL accuracy in parton showers*, *JHEP* **03** (2023) 224 [2301.09645].

[9] P. Sarmah, A. Siódmok and J. Whitehead, *KrkNLO matching for colour-singlet processes*, *JHEP* **01** (2025) 062 [2409.16417].

[10] P. Sarmah, A. Siódmok and J. Whitehead, *KrkNLO matching and phenomenology for vector boson processes*, 2511.16605.

[11] S. Jadach, W. Płaczek, S. Sapeta, A. Siodmok and M. Skrzypek, *Parton distribution functions in Monte Carlo factorisation scheme*, *Eur. Phys. J. C* **76** (2016) 649 [1606.00355].

[12] W.A. Bardeen, A.J. Buras, D.W. Duke and T. Muta, *Deep Inelastic Scattering Beyond the Leading Order in Asymptotically Free Gauge Theories*, *Phys. Rev. D* **18** (1978) 3998.

[13] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[14] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650.

[15] T.D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, *Phys. Rev.* **133** (1964) B1549.

[16] S. Frixione, Z. Kunszt and A. Signer, *Three jet cross-sections to next-to-leading order*, *Nucl. Phys. B* **467** (1996) 399 [hep-ph/9512328].

[17] V.V. Sudakov, *Vertex Parts at Very High Energies in Quantum Electrodynamics*, *Sov. Phys. JETP* **3,4,5,6** (1956) 65.

[18] S. Frixione, P. Nason and B.R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, *JHEP* **08** (2003) 007 [hep-ph/0305252].

[19] HEP Software Foundation collaboration, *HL-LHC Computing Review: Common Tools and Community Software*, in *Snowmass 2021*, P. Canal et al., eds., 8, 2020, DOI [2008.13636].

[20] P. Nason, *MINT: A Computer program for adaptive Monte Carlo integration and generation of unweighted distributions*, 0709.2085.

[21] R. Frederix and P. Torrielli, *A new way of reducing negative weights in MC@NLO*, *Eur. Phys. J. C* **83** (2023) 1051 [2310.04160].

[22] Y. Che, "Reducing Negative Weights in MC@NLO by Improved Implementation of Born Spreading." Bachelor Thesis, http://lup.lub.lu.se/student-papers/record/9157443, 2024.

[23] S.E. Farkh, R. Frederix and M. Gouighri, *MC@NLO event generation by reweighting unweighted Born events*, 2602.18124.

[24] R. Frederix, S. Frixione, S. Prestel and P. Torrielli, *On the reduction of negative weights in MC@NLO-type matching procedures*, *JHEP* **07** (2020) 238 [2002.12716].

[25] A. Buckley, P. Ilten, D. Konstantinov, L. Lönnblad, J. Monk, W. Pokorski et al., *The HepMC3 event record library for Monte Carlo event generators*, *Comput. Phys. Commun.* **260** (2021) 107310 [1912.08005].

[26] H. Dembinski, E. Rodrigues, H. Schreiner, A. Desai, A. Prosekin, D. Kobylianskii et al., *scikit-hep/pyhepmc: 2.16.1*, Dec., 2025. 10.5281/zenodo.17979895.

[27] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021) 014013 [1912.10053].

[28] M. van Beekveld, S. Ferrario Ravasio, J. Helliwell, A. Karlberg, G.P. Salam, L. Scyboz et al., *Logarithmically-accurate and positive-definite NLO shower matching*, *JHEP* **10** (2025) 038 [2504.05377].

[29] S. Catani and M.H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys. B* **485** (1997) 291 [hep-ph/9605323].

[30] S. Delorme, A. Kusina, A. Siódmok and J. Whitehead, *Factorisation schemes for proton PDFs*, *Eur. Phys. J. C* **85** (2025) 505 [2501.18289].

[31] H. Wiegand, *Kish, l.: Survey sampling. john wiley & sons, inc., new york, london 1965, ix + 643 s., 31 abb., 56 tab., preis 83 s.*, *Biometrische Zeitschrift* **10** (1968) 88 [https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.19680100122].