Color-singlet production at NNLO in MCFM

Radja Boughezal\textsuperscript{1,a}, John M. Campbell\textsuperscript{2,b}, R. Keith Ellis\textsuperscript{3,c}, Christfried Focke\textsuperscript{4,d}, Walter Giele\textsuperscript{3,e}, Xiaohui Liu\textsuperscript{5,f}, Frank Petriello\textsuperscript{1,4,g}, Ciaran Williams\textsuperscript{6,h}

\textsuperscript{1} High Energy Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA
\textsuperscript{2} Fermilab, P.O.Box 500, Batavia, IL 60510, USA
\textsuperscript{3} Department of Physics, Institute for Particle Physics Phenomenology, University of Durham, Durham DH1 3LE, UK
\textsuperscript{4} Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA
\textsuperscript{5} Maryland Center for Fundamental Physics, University of Maryland, College Park, Maryland 20742, USA
\textsuperscript{6} Department of Physics, University at Buffalo, The State University of New York, Buffalo 14260, USA

Received: 14 July 2016 / Accepted: 7 December 2016 / Published online: 30 December 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We present the implementation of several color-singlet final-state processes at Next-to-Next-to Leading Order (NNLO) accuracy in QCD to the publicly available parton-level Monte Carlo program MCFM. Specifically we discuss the processes $pp \rightarrow H$, $pp \rightarrow Z$, $pp \rightarrow W$, $pp \rightarrow Hz$, $pp \rightarrow HW$ and $pp \rightarrow \gamma \gamma$. Decays of the unstable bosons are fully included, resulting in a flexible fully differential Monte Carlo code. The NNLO corrections have been calculated using the non-local $N$-jettiness subtraction approach. Special attention is given to the numerical aspects of running MCFM for these processes at this order. We pay particular attention to the systematic uncertainties due to the power corrections induced by the $N$-jettiness regularization scheme and the evaluation time needed to run the hybrid openMP/MPI version of MCFM at NNLO on multi-processor systems.

1 Introduction

The second run of the LHC (Run II) which is currently under way, will result in the accumulation of an unprecedented amount of high-quality data in a new high energy regime. In tandem with the well-understood and carefully calibrated detectors, this will lead to experimental uncertainties that are at the level of a few percent or smaller for many of the most important processes. These include various Higgs boson production channels, as well as standard candle processes such as vector boson production. Studies of diboson production will allow for stringent tests of the Electroweak sector of the Standard Model (SM) and constraints on possible new physics scenarios. In order to make best use of the precise experimental observations it is crucial to have access to accurate theoretical calculations of the same quantities. At the LHC this requires the calculation of QCD corrections to inclusive and differential cross sections at increasingly higher order. For the most efficient comparison between theoretical predictions and experimental data it is extremely beneficial for theoretical results to be released in the form of a public code, allowing users full flexibility in obtaining theoretical predictions relevant for their analysis.

While calculations at next-to-leading order (NLO) in the strong coupling constant are by now quite standard, only about 20 processes have been calculated through to next-to-next-to-leading order (NNLO). Recent publications on these processes are shown in Table 1.

All such calculations require a means by which to regulate the soft and collinear radiation that appears in the calculation of the higher-order contributions. At NLO local subtraction schemes, such as FKS\textsuperscript{41} or Catani–Seymour dipole subtraction\textsuperscript{42}, are typically preferred. In these local subtraction formalisms, the singular unresolved infrared limits are cancelled point-wise by local counterterms. These local counterterms, after analytic integration over the unresolved partons, are added to the virtual corrections yielding a finite result.
The construction of a local subtraction scheme for a NNLO calculation is a daunting task, given the complexities of the multiple infrared limits and differing dimensionality of phase space for the component parts. However, progress has been made, with significant advances over the last decade. The first local subtraction scheme used at NNLO was based on the sector decomposition approach that had previously been applied to multi-loop integrals [43,44] and was later adapted to the case of phase-space integrals [45,46]. This scheme separates the overlapping singularities by using a plus-prescription to isolate the singular contributions, thereby avoiding any analytic integrations over regions of phase space. The antenna subtraction method was extended to NNLO in Refs. [47,48], and has been used to obtain predictions for 2 → 2 processes in which both final-state particles are colored [40]. Antenna subtraction resembles the NLO subtraction formalisms in that the doubly unresolved limits are cancelled point-by-point in phase space by counterterms which require analytic integration to cancel infrared poles in the real-virtual and double virtual phase spaces. Finally in Refs. [49–51] the sector decomposition approach was generalized to arbitrary processes. By partitioning the phase space into appropriate sectors in which each singularity can be made manifest, and then performing a Laurent series expansion to extract the poles. This method has been applied to various processes at the LHC [5,7,34,38].

In addition to the local subtraction schemes discussed above, there is an alternate form of regulation, which is inherently non-local. Indeed one of the first NLO regularization techniques developed was one such method, phase-space slicing, introduced in Refs. [52–54]. In these methods a parameter is used to separate the resolved and unresolved phase spaces. The resolved region of phase space corresponds to a calculation of the process with one additional final-state parton, and if a suitable resolution parameter is chosen, the unresolved region can be directly calculated. At NLO non-local methods have generally fallen out of favor. This is due to the large cancellation between the resolved and unresolved contributions at small values of the resolution parameter, which can induce large Monte Carlo uncertainties.

However, non-local subtraction schemes have made a resurgence for NNLO calculations. Although they have the disadvantages discussed above they also have several advantages which make them attractive for NNLO calculations. First, they are conceptually simple to implement. Once a suitable resolution parameter is selected, the singly unresolved part of the calculation can be obtained with existing NLO event generators, such as MCFM [55–57]. Second, with recent advances in computing, the drawback associated with the large numerical cancellations can be mitigated by running with a large number of computer cores. Finally by using a resolution parameter motivated by a physical factorization theorem, the approximations inherent in the method can be systematically improved, e.g. by analytic calculations of power-suppressed contributions [58].

The first non-local subtraction developed for NNLO calculation was the so-called $q_T$ subtraction method [3]. This method uses the transverse momentum of the final-state color neutral particle, $q_T$, as the cut variable. For $q_T < q_T^{\text{cut}}$ the factorization theorem of Collins, Soper and Sterman [59], can be used to compute the cross section, while above the cutoff the NLO calculation of the color singlet plus jet can be utilized. An obvious drawback is that it is only applicable to color neutral final states. Inspired by a factorization formula [60] from soft collinear effective field theory (SCET) [61–65] the first steps toward extending these ideas to calculations containing colored final states were taken in a calculation of top-quark decay at NNLO [37]. However, no initial collinear singularities appear in this calculation. A powerful generalization of this idea applicable to general initial and final states was introduced in [4,16]. It is obtained by replacing the $q_T$ variable with the event-shape $N$-jettiness variable [66]. Below the $N$-jettiness ($\tau_N$) cutoff, SCET provides the relevant factorization theorem [66]. For the below-cut region the necessary SCET ingredients, corresponding to the final-state and initial-state collinear radiation functions are already known, and are represented by the two-loop jet-functions of Refs. [67,68] and the two-loop beam function of Ref. [69,70]. The corresponding two-loop soft functions are also known for zero-jettiness [71,72] and for general $N$-jettiness [73]. The first process calculated at NNLO using this method was $pp \to W + \text{jet}$ [16], followed by calculations of the $pp \to H + \text{jet}$ [8] and $pp \to Z + \text{jet}$ [18] processes, and by detailed phenomenological studies of these processes at this order [74–76]. $pp \to H$ and $pp \to Z$ were also calculated using this method [4]. Additional processes of phenomenological interest, $pp \to VH$ [22] and $pp \to \gamma \gamma$ [33] have been calculated using the same approach.

As mentioned before, an important advantage of the $N$-jettiness subtraction method is that it meshes well with the

---

**Table 1** Publications on processes evaluated differentially at NNLO

| Process | Reference |
|---------|-----------|
| $H + 0\ \text{jet}$ | [1–4] |
| $W + 0\ \text{jet}$ | [13,14] |
| $ZH$ | [21,22] |
| $ZZ$ | [25–27] |
| $t\bar{t}$ | [34,35] |
| $H + 1\ \text{jet}$ | [5–9] |
| $Z/\gamma^* + 0\ \text{jet}$ | [4,14,15] |
| $WH$ | [22,23] |
| $WW$ | [28–30] |
| Single top | [36] |
| $W + 1\ \text{jet}$ | [16] |
| $W + \gamma, Z + \gamma$ | [31] |
| $WZ$ | [24] |
| $\gamma\gamma$ | [32,33] |
| $H \to b\bar{b}$ | [11,12] |
| $VH$ | [22] |
| $VH$ | [31] |
| Top decay | [37,38] |
| $1\ \text{jet}$ | [16] |
| $1\ \text{jet}$ | [16] |
| $1\ \text{jet}$ | [16] |
| $2\ \text{jet}$ | [16] |
| $2\ \text{jet}$ | [16] |
| $2\ \text{jet}$ | [16] |
| $H → b\bar{b}$ | [11,12] |
| $H → b\bar{b}$ | [11,12] |
| $H → b\bar{b}$ | [11,12] |
| $H → b\bar{b}$ | [11,12] |
existing NLO calculations, such as those included in MCFM. Included in MCFM are the NLO corrections to \( W + n \) jet production, \( Z + n \) jets production \([77]\), Higgs \( +n \) jets production \([78, 79]\) for \( n = 0, 1, 2 \), making the implementation of \( W, Z, H + 0.1 \) jet at NNLO possible.

The recent advances in NNLO technologies allows for the exciting possibility of releasing a public code capable of computing many \( 2 \rightarrow 2 \) processes at NNLO accuracy. This paper presents a first step in this journey by summarizing the implementation of the \( N \)-jettiness subtraction procedure in MCFM, and presenting a detailed breakdown of the method for the processes released in the initial version of the NNLO code. An important consideration in making the code public is computational speed. In Ref. \([57]\) MCFM was upgraded to use a parallel version of the VEGAS adaptive integration method using openMP. For NNLO calculations, this was expanded by using a hybrid g/mpi version of MCFM for use on computing clusters to facilitate the numerical NNLO calculations of Ref. \([18]\). Using the hybrid version of MCFM we can calculate NNLO distributions efficiently within a reasonable timescale.

In summary, this paper describes the implementation of the \( N \)-jettiness subtraction procedure in MCFM and presents results for the processes available in MCFM v8.0. Specifically these processes are \( pp \rightarrow H, W, Z, VH, \gamma \gamma \). Where present, the decays of unstable particles are included, allowing for a fully flexible MC code. In Sect. 2 we will give a schematic overview of the non-local \( N \)-jettiness subtraction scheme. Section 3 will present details of the calculational set-up and Sects. 4 and 5 will look at the results. In Sect. 6 processes are summarised. Finally in Sect. 7 the main results are summarized.

## 2 SCET-based non-local subtraction

A collision of partons \( a \) and \( b \) with momentum fractions \( x_a, b \), originating from the incoming beam protons with momenta \( p_{a,b} \), produces a final state including \( N \) jets with momenta \( \{ q_i \} \). The jettiness of parton \( j \) with momentum \( p_j \) is defined as

\[
T_N(p_j) = \min_{i=a,b,1,...,N} \left\{ \frac{2 q_i \cdot p_j}{Q_i} \right\},
\]

where for notational simplicity we have set \( q_{a,b} = p_{a,b} \). We denote the jet or beam energy by \( E_j \), \( Q_i \) is a measure of the jet/beam hardness. In our numerical results we set this equal to twice the jet/beam energy, \( Q_i = 2E_i \) \([66]\). We can now define the event jettiness, or \( N \)-jettiness, as the sum over all the \( M \) final-state parton jettiness values

\[
T_N = \sum_{k=1}^{M} T_N(p_k) = \sum_{k=1}^{M} \min_{i=a,b,1,...,N} \left\{ \frac{2 q_i \cdot p_k}{Q_i} \right\}.
\]

For Leading Order (LO) events we have \( \{ p_i \} = \{ q_i \} \) and the event jettiness is zero. Beyond LO (\( M > N \)), only in the soft/collinear limit will the event jettiness necessarily go to zero. Therefore the event \( N \)-jettiness can be used in a non-local subtraction approach where we can isolate the doubly unresolved region of the phase space by demanding \( T_N < T_N^{\text{cut}} \). In this paper we restrict ourselves to color-singlet final-state events. We can therefore use the event-shape variable \( T_0 \) to regulate the initial-state radiation.

By demanding \( T_0 < T_0^{\text{cut}} \) one isolates the doubly unresolved regions of phase space. The matrix elements in the soft/collinear approximation can be analytically integrated over this region and added to the virtual contributions. The regions of phase space where \( T_0 > T_0^{\text{cut}} \) are integrated over numerically. In the limit \( T_0^{\text{cut}} \rightarrow 0 \) this will result in the correct results for the cross section.

To obtain the analytic soft/collinear expressions we use all-orders resummation results which rely heavily on the machinery of soft-collinear effective theory (SCET) \([61–65]\). The all-orders resummation of the \( T_0 \) event-shape variable in the limit \( T_0 \rightarrow 0 \) was constructed in Ref. \([66]\):

\[
\frac{d\sigma}{dT_0} = \sum_{ab} \int dx_a dx_b \int d\Phi_B(p_a, p_b; \Phi_{\text{singlet}}) \Theta(\Phi_{\text{singlet}})
\times H_{ab}(\Phi_B, \mu) \frac{d\Delta_{ab}}{dT_0} + \cdots,
\]

where the indices \( a, b \) run over all initial-state partons involved in the scattering. The initial-state momenta \( p_{a,b} \) are given by the momenta fractions \( x_{a,b} \), while \( \Phi_B \) denotes the Born-level color-singlet phase space \( p_a p_b \rightarrow \Phi_{\text{singlet}} \). The composite \( \Theta(\Phi_{\text{singlet}}) \) denotes any phase-space restrictions on the color-singlet phase space. The soft/collinear function \( \Delta_{ab} \) is given by

\[
\frac{d\Delta_{ab}}{dT_0} = B_{a} \otimes B_b \otimes S_{ab}
\]

\[
\equiv \int dt_{B_a} dt_{B_b} dt_{S} \delta(t_0 - t_{B_a} - t_{B_b} - t_S) \times B_{a}(t_{B_a}, x_a, \mu) B_b(t_{B_b}, x_b, \mu) S_{ab}(t_S, \mu).
\]

A summary of the various components which appear in these expressions is given below:

- The hard function \( H \) encodes the effect of hard virtual corrections. At leading order in the \( \alpha_s \)-expansion it reduces to the leading-order partonic cross section. At higher orders it also contains the finite contributions of the pure virtual corrections, renormalized at scale \( \mu \) using the \( \overline{\text{MS}} \) scheme. It depends on the Born-level kinematics and the scale choice.
The beam function $B_{a}$ contains the effects of initial-state collinear radiation. It depends on $t_{B_{a}}$, the contribution of initial-state collinear radiation to $T_{0}$. The beam function is non-perturbative; however, up to corrections suppressed by $\Delta_{QCD}/t_{B}$, it can be written as a convolution of perturbative matching coefficients and the usual parton density functions, $f_{i}/H$, 

$$B_{a}(t_{B_{a}}, x, \mu) = \sum_{i} \int \frac{d\xi}{\xi} T_{a\xi}(t_{B_{a}}, x/\xi, \mu) f_{i\xi}(\xi),$$

(2.5)

where we have suppressed the scale dependence of the parton density functions, and $i$ runs over all partons. The two-loop beam functions have been computed in Refs. [69,70].

The soft function $S$ collects the jetiness contributions of soft radiation. It depends on $t_{S}$, the contribution of soft radiation to $T_{0}$. The expansion of the soft function for zero-jetness up to two-loop order can be found in Refs. [71,72].

The delta function appearing in Eq. (2.4) combines the contribution of each type of radiation to produce the measured value of $T_{0}$. The factorization formula is correct up to power corrections, indicated by the ellipsis in Eq. (2.3). These power corrections can in principle be calculated in the same way as one derives the leading power components in Eq. (2.4). However, they can be neglected as long as we restrict ourselves to the phase-space region $T_{0} \ll Q$, where $Q$ denotes the hard scale in the process (for the zero-jet processes considered here, $Q$ is of the order of the invariant mass of the final state). Integrating Eq. (2.3) over the region $T_{0} < T_{0}^{\text{cut}}$ will give the analytic result for the below-cut cross section:

$$d\sigma(T_{0}^{\text{cut}}) = \sum_{ab} \int dx_{a} dx_{b} \int d\Phi_{B}(p_{a}, p_{b}; p_{\text{singlet}}) \Theta(p_{\text{singlet}}) \times H_{ab}(\Phi_{B}, \mu) \Delta_{ab}(T_{0}^{\text{cut}}) + \cdots,$$

(2.6)

with

$$\Delta_{ab}(T_{0} < T_{0}^{\text{cut}}) = \int_{0}^{T_{0}^{\text{cut}}} dT_{0} (B_{a} \otimes B_{b} \otimes S_{ab}).$$

(2.7)

Next we expand the functions order by order in $\alpha_{S}$ using a superscript to denote the power of $\alpha_{S}$ appearing in each term. That is, we expand any perturbative function $\mathcal{F}$ as

$$\mathcal{F} = \mathcal{F}^{(0)} + \alpha_{S} \mathcal{F}^{(1)} + \alpha_{S}^{2} \mathcal{F}^{(2)} + \cdots.$$

(2.8)

This results in

$$d\sigma^{(n)}(T_{0}^{\text{cut}}) = \sum_{ab} \int dx_{a} dx_{b} \int d\Phi_{B}(p_{a}, p_{b}; p_{\text{singlet}}) \Theta(p_{\text{singlet}}) \times \sum_{k=0}^{n} H_{ab}(\Phi_{B}, \mu) \Delta_{ab}^{(k)}(T_{0}^{\text{cut}}),$$

(2.9)

$$\Delta_{ab}^{(n)}(T_{0}^{\text{cut}}) = \sum_{k+l+m=n} \int_{0}^{T_{0}^{\text{cut}}} dT_{0} \left( B_{a}^{(k)} \otimes B_{b}^{(l)} \otimes S_{ab}^{(m)} \right).$$

To obtain the $O(\alpha_{S}^{2})$ correction to the soft/collinear cross section below the $T_{0}^{\text{cut}}$ we need

$$\Delta_{ab}^{(0)}(T_{0}^{\text{cut}}) = \int_{0}^{T_{0}^{\text{cut}}} dT_{0} \left( B_{a}^{(0)} \otimes B_{b}^{(0)} \otimes S_{ab}^{(0)} \right),$$

(2.10)

$$\Delta_{ab}^{(1)}(T_{0}^{\text{cut}}) = \int_{0}^{T_{0}^{\text{cut}}} dT_{0} \left( B_{a}^{(1)} \otimes B_{b}^{(0)} \otimes S_{ab}^{(0)} + B_{b}^{(0)} \otimes B_{a}^{(1)} \otimes S_{ab}^{(0)} + B_{a}^{(0)} \otimes B_{b}^{(0)} \otimes S_{ab}^{(1)} \right),$$

(2.11)

Note that the leading-order expressions for the subtraction functions are proportional to delta functions in their respective hadronic variable:

$$B_{a}^{(0)}(t_{B_{a}}, x_{a}, \mu) = \delta(t_{B_{a}}) f_{a}/H(x_{a}); S_{ab}^{(0)}(t_{s}, \mu) = \delta(t_{s}).$$

(2.12)

The soft and beam function have the generic forms

$$S^{(n)}(t_{s}, \mu) = s_{-1}^{(n)} \delta(t_{s}) + \sum_{k=0}^{n-1} s_{k}^{(n)} L_{k}(t_{s}, \mu),$$

(2.13)

$$T_{ij}^{(n)}(t_{a}, z, \mu) = t_{-1,ij}(z) \delta(t_{a}) + \sum_{k=0}^{n-1} t_{k,ij}(z) L_{k}(t_{a}, \mu),$$

(2.14)

$$L_{n}(t, \mu) = \frac{1}{\mu} \left[ \frac{\mu}{\mu} \ln^{n}(t) \right].$$

(2.15)

For example, the following contributions to the NNLO SCET function become

$$\int_{0}^{T_{0}^{\text{cut}}} dT_{0} S^{(2)} \otimes T_{ai}^{(0)} \otimes T_{bj}^{(0)}$$

$$= \delta_{ai} \delta_{bj} \left\{ s_{-1}^{(2)} + \sum_{n=0}^{3} \frac{1}{n+1} s_{n}^{(2)} L^{n+1} \right\},$$
\[
\int_0^{T_0^{\text{cut}}} \mathcal{L}_0 \, S^{(1)} \otimes T_{ai}^{(1)} \otimes T_{bj}^{(0)} \\
= \delta_{bj} \left\{ s_{-1}^{(1)} t_{-1,ai}(z) + s_{-1}^{(1)} \sum_{n=0}^{1} \frac{1}{n+1} i_{n,ai}^{(1)}(z) L^{n+1} \\
+ i_{-1,ai}^{(1)}(z) \sum_{n=0}^{1} \frac{1}{n+1} s_{n}^{(1)} L^{n+1} \\
+ \frac{1}{m,n=0} s_{m}^{(1)} i_{n,ai}^{(1)}(z) \Gamma_{m,n} \right\}, \quad (2.13)
\]

where

\[
L = \ln \left( \frac{T_0^{\text{cut}}}{\mu} \right). \quad (2.14)
\]

and

\[
\begin{align*}
\Gamma_{0,0} &= L^2 - \xi_2 \\
\Gamma_{1,0} &= \Gamma_{0,1} = \frac{L^3}{2} - \xi_2 L + \xi_3 \\
\Gamma_{1,1} &= \frac{L^4}{4} - \xi_2 L^2 + 2\xi_3 L - \frac{\xi_2^2}{10}.
\end{align*} \quad (2.15)
\]

Using these results it is possible to analytically compute all of the necessary hadronic integrals in Eq. (2.10). The remaining integrals are over the Born phase space and parton distribution functions, and are simple to perform numerically. This completes the calculation of the \( T < T_0^{\text{cut}} \) phase-space region. We note that the cross section below \( T_0^{\text{cut}} \) will contain terms of the form \( \ln^n(T_0^{\text{cut}}) \), where \( n \) ranges from 0 to 4 at NNLO. An important check of our framework is the cancellation of these terms against the identical logarithms that appear for \( T_0 > T_0^{\text{cut}} \). We must also choose \( T_0^{\text{cut}} \) small enough to suppress the power corrections in Eq. (2.3). Both of these issues will be addressed in the subsequent sections.

### 3 Process overview

For all of the studies performed in this paper we perform calculations for the LHC operating at a center-of-mass energy of \( \sqrt{s} = 13 \text{ TeV} \). The parameters that are used throughout this paper are shown in Table 2. Finally, we use the NNLO MSTW2008 PDF set (MSTW8nnm) that corresponds to \( \alpha_s(M_Z) = 0.11707 \).

An overview of the processes that will be studied in detail in this paper is shown in Table 3.\(^1\) As well as detailing the default choice of renormalization and factorization scales (\( \mu_R \) and \( \mu_F \)), this table also shows the corresponding cross section up to NNLO. The NNLO cross sections are written in the form

\[
\sigma_{\text{NNLO}} = \sigma_{\text{LO}} \times \left( 1 + \frac{\Delta\sigma_{\text{LO}}}{\sigma_{\text{LO}}} + \frac{\Delta\sigma_{\text{NNLO}}}{\sigma_{\text{LO}}} \right), \quad (3.1)
\]

so that, for instance, the corresponding NLO result is obtained by simply omitting the final term in this equation. The cross sections have been obtained by running the readily available public codes referenced in the final column of Table 3.

We now describe the calculational setup that we use for these processes, which corresponds to the default behavior of the above codes. This behavior has been matched in the MCFM code and, in order to establish the equivalence of the parameters for MCFM and the other publicly available codes, we compare results up to NLO in Table 4. The agreement is excellent for all processes, so that we can be sure that MCFM should produce the same results as the other codes when computing the NLO and NNLO predictions using the \( N \)-jettiness subtraction method.

#### 3.1 Higgs production through gluon fusion

We work in a theory in which only the top quark has a non-zero Yukawa coupling. Taking the large \( m_t \) limit we obtain an effective Lagrangian that expresses the coupling of gluons to the Higgs field \cite{84},

\[
\mathcal{L}^\text{int}_H = \frac{C(m_t^2, \mu^2)}{2} H \sum_a G^a_{\mu\nu} G^{a\mu\nu}. \quad (3.2)
\]

where the sum is over the color degrees of freedom of the gluon. At the order required in this paper, the coefficient \( C(m_t^2, \mu^2) \) is given in the MS scheme by \cite{85,86},

\[
C(m_t^2, \mu^2) = \frac{\alpha_s}{6\pi v} \left[ 1 + \frac{\alpha_s}{4\pi} (5C_A - 3C_F) \right] + \left( \frac{\alpha_s}{4\pi} \right)^2 \left[ \frac{27}{2} C_F C_A + \left( 11 \ln \frac{m_t^2}{\mu^2} - \frac{100}{3} \right) C_F C_A \right]
\]
To enhance the size of the NNLO correction we therefore perform a careful study of the behavior of the subtraction method with the codes used for cross-checking in the final column for processes considered in this paper. NLO and NNLO corrections are shown as relative enhancements to the LO cross section. $q^2$ is the overall invariant mass squared of the vector boson and the Higgs boson.

| Process          | $\mu_R$ | $\mu_F$ | Cross section to NNLO | Reference       |
|------------------|---------|---------|-----------------------|-----------------|
| $g g \rightarrow H$ | $M_H$   | $M_H$   | $12.937 \times (1 + 1.28 + 0.77)$ pb | ggh@nnlo [80]   |
| $Z$              | $2M_Z$  | $M_Z/2$ | $44.303 \times (1 + 0.22 + 0.05)$ nb | ZWMS [81]       |
| $W^+$            | $2M_W$  | $M_W/2$ | $81.561 \times (1 + 0.23 + 0.06)$ nb | ZWMS [81]       |
| $ZH$             | $\sqrt{q^2}$ | $\sqrt{q^2}$ | $0.68255 \times (1 + 0.16 + 0.10)$ pb | vh@nnlo [82,83] |
| $W^+H + W^-H$    | $\sqrt{q^2}$ | $\sqrt{q^2}$ | $1.2593 \times (1 + 0.16 + 0.02)$ pb | vh@nnlo [82,83] |

Table 4 Comparison of LO and NLO cross sections computed using the standard MCFM subtraction method with the codes used for cross-checking in this paper.

| Process | Order | MCFM cross section | Cross-check |
|---------|-------|---------------------|-------------|
| $H$ production | LO | $12.937 \pm 0.001$ pb | $12.937$ pb |
|          | NLO  | $29.520 \pm 0.001$ pb | $29.521 \pm 0.001$ pb |
| $Z$     | LO   | $44.303 \pm 0.001$ nb | $44.303$ nb |
|          | NLO  | $53.958 \pm 0.002$ nb | $53.957 \pm 0.001$ nb |
| $W^+$   | LO   | $81.561 \pm 0.002$ nb | $81.561$ nb |
|          | NLO  | $100.298 \pm 0.003$ nb | $100.299 \pm 0.001$ nb |
| $ZH$    | LO   | $0.68254 \pm 0.00001$ pb | $0.68255$ pb |
|          | NLO  | $0.79073 \pm 0.00003$ pb | $0.79079 \pm 0.00006$ pb |
| $W^+H + W^-H$ | LO | $1.2592 \pm 0.00001$ pb | $1.2593$ pb |
|          | NLO  | $1.4629 \pm 0.00001$ pb | $1.4630 \pm 0.00001$ pb |

\[
- \left(7 \frac{m_t^2}{\mu^2} - \frac{1063}{36}\right) C_A^2 \\
- \frac{4}{3} C_F T_F - \frac{5}{6} C_A T_F - \left(8 \ln \frac{m_t^2}{\mu^2} + 5\right) C_F T_F n_f \\
- \frac{47}{9} C_A T_F n_f \right) \\
\]

(3.3)

Here $\mu$ is the vacuum expectation value of the Higgs field, $\mu = 246$ GeV. The only remaining $m_t$-dependence at this order is the one shown in the $O(\alpha_S^3)$ contribution to the coefficient of the effective operator.

The validation cross section for this process is obtained using ggh@nnlo [80]. As can be seen from Table 3, at 13 TeV the higher-order corrections to this cross section are quite large.

3.2 $W$ and $Z$ production

To establish the correct values of the higher-order cross sections for $W^+$ and $Z$ production we use the program ZWMS [81]. For the sake of illustration we have chosen to perform the comparison for only one charge of the $W$-boson. We note that for the canonical scale choice $\mu_R = \mu_F = M_V$ (where $V = W, Z$) the NNLO corrections are very small. Although this is ultimately an advantage in terms of the accuracy required for phenomenological applications, it prohibits a careful study of the behavior of the $N$-jettiness calculation. To enhance the size of the NNLO correction we therefore use an asymmetric choice, $\mu_R = 2M_V$, $\mu_F = M_V/2$. This results in NNLO corrections of approximately 5% relative to the LO cross section at 13 TeV (cf. Table 3).

Note that by default MCFM includes the decay of the vector bosons, $Z/\gamma^*$ to a lepton pair. For comparison with the rate of production for on-shell $Z$-bosons, we remove the (small) contribution mediated by a virtual photon and divide out the overall branching ratio of the $Z$-boson to leptons.

3.3 Associated Higgs production: $W^\pm H$ and $ZH$ processes

To establish target NLO and NNLO values for the total cross section for $W^\pm H$ and $ZH$ production we use the program vh@nnlo [82,83]. In order to facilitate an easy comparison with this program, we use the scale choice $\mu_R = \mu_F = \sqrt{q^2} \equiv \sqrt{(p_V + p_H)^2}$, with $V = W^\pm$ or $V = Z$ as appropriate. We also sum over both charges of the $W$ boson, i.e. we include both $W^+$ and $W^-$ contributions—which can differ substantially at a $pp$ collider such as the LHC—in all of the results below. For the diagrams in which the Higgs boson couples directly to a top-quark loop we work in the effective theory, valid in the large $m_t$ limit given by Eq. (3.2). A detailed phenomenological study of the NNLO implementation of these processes in MCFM has been presented in Ref. [22].

For both $W^\pm H$ and $ZH$ processes the correction originating from diagrams with the Higgs boson coupling to a top-quark loop is approximately 1.5%. The $ZH$ process also...
includes a substantial finite component due to \( gg \rightarrowZH \) loops at NNLO. The NNLO corrections that correspond to simple dressings of the LO diagrams are very small, of order 1%, for both \( W^\pm H \) and \(ZH \) production. The net effect of all these contributions is shown in Table 3, where the NNLO corrections to the \(ZH \) process are at the level of 10%. In contrast, the total NNLO correction to the \( W^\pm H \) process is about 2% of the LO result.

### 3.4 Diphoton production

NNLO predictions for the diphoton process, obtained using MCFM, have been presented in Ref. [33]. Therein the results have been validated using the same procedure as we will adopt later; we do not repeat that analysis here. However, we will later on summarize the size of the power corrections and timing results for this process.

### 4 N-Jettness subtraction at NLO

Although the calculation of NLO corrections for the processes considered here is straightforward, a detailed examination of the corresponding N-jettness subtraction calculation is extremely useful. It provides a stringent check of the accuracy of this approach, namely a direct probe of the size of the power corrections that have been neglected in Eq. (2.3). This can be tested with exquisite accuracy, due to the relative simplicity of the calculation compared to the corresponding exercise at NNLO. This comparison can also illuminate the limitations of this approach when moving beyond an inclusive calculation, by using MCFM to compare calculations of differential distributions at NLO.

The calculation of NLO corrections using the N-jettness subtraction method is straightforward in MCFM. The below-cut contribution is easily computed, while the above-cut contribution corresponds to a LO calculation of the process that contains an additional parton. In order to avoid numerical instability in calculations using MCFM, previous versions of the code have applied a small cutoff on all invariant masses present in the problem, \( \sqrt{s_{ij}} > c\text{utoff} \). In this version this has been changed so as to enforce a small cutoff on the partonic jettiness of every parton present in a given calculation, \( T_n(p_j) > c\text{utoff} \). Since the above-cut region involves a standard LO calculation, for which there are no numerical instabilities, we are able to choose a value for this cutoff close to the limit of double precision, \( c\text{utoff} = 10^{-12} \text{ GeV} \).

#### 4.1 Inclusive cross sections

The benchmark cross sections that form the basis for this comparison can be extracted from Table 4 and, for convenience, have been summarized in Table 5. As is well-known, the NLO corrections to Higgs production through gluon fusion are very large, while all of the other processes receive corrections of order 20%.

A comparison of the N-jettness calculations of these coefficients, with the results shown in Table 5, is shown in Fig. 1. The ratio of the calculations is shown as a function of \( T_{0}^{\text{cut}} \), for a range of suitable values of \( T_{0}^{\text{cut}} \). The approach of the N-jettness calculation to the correct result as \( T_{0}^{\text{cut}} \rightarrow 0 \) is clear for each process. However, the manner in which the correct result is reached varies considerably. For instance, Higgs production through gluon fusion approaches the correct result from above, while the other processes approach it from below. The approach is much slower for \( W^+ \) and \( Z \) production than for any of the Higgs production processes, with percent level accuracy only reached for \( T_{0}^{\text{cut}} \lesssim 0.01 \text{ GeV} \).

The approach of the N-jettness subtraction result to the correct answer is determined by the behavior of power corrections that are not accounted for at present. At the NLO level after integration over the final-state phase space this can be modeled by the following functional form:

\[
\Delta\sigma_{\text{jettness}}^{\text{NLO}}(T_{0}^{\text{cut}}) = \Delta\sigma^{\text{NLO}} + c \times \left( \frac{T_{0}^{\text{cut}}}{Q} \right) \times \log \left( \frac{T_{0}^{\text{cut}}}{Q} \right),
\]

(4.1)

where \( Q \) is the appropriate scale for the process at hand and \( c \) is an unknown constant. For single boson production \( Q \) is taken to be the mass of the produced particle \( (M_H, M_W \text{ or } M_Z) \) while for the associated production processes we use \( Q = M_W + M_H \) and \( Q = M_Z + M_H \). Figure 1 also shows a fit of the results to Eq. (4.1), with the values of \( \Delta\sigma^{\text{NLO}} \) and \( c \) determined in the fit. The difference of the fit value for \( \Delta\sigma^{\text{NLO}} \) with the known result given in Table 5 is no larger than one per mille for all processes.

Since the speed of the approach to the correct result is qualitatively much worse for \( W \) and \( Z \) production it is instructive to examine the processes in more detail in order to uncover the origin of the difference. To that end we now turn to a comparison of more differential results.

| Process | \( \Delta\sigma^{\text{NLO}} \) | \( \Delta\sigma^{\text{NLO}}/\sigma^{\text{LO}} \) |
|---------|-------------------------------|-----------------------------------------------|
| \( H \) | 16.58 pb | 1.282 |
| \( Z \) | 9.655 nb | 0.218 |
| \( W^+ \) | 18.74 nb | 0.230 |
| \( ZH \) | 0.1082 pb | 0.158 |
| \( W^+H + W^-H \) | 0.2037 pb | 0.162 |
4.2 Rapidity distributions at NLO

The simplest distribution to study is the rapidity of the produced system, which is intimately related to the momentum fractions carried by the incident partons. We will compare the prediction for the NLO contribution to this distribution (i.e. corresponding to $\Delta y_{\text{NLO}}$) computed using dipole subtraction and jettiness subtraction with $T_0^\text{cut} = 0.01$ GeV and $T_0^\text{cut} = 0.04$ GeV. The difference between the true result and the jettiness calculation for $T_0^\text{cut} = 0.04$ GeV is about 0.4% for $gg \rightarrow H$, 1.5% for $Z$ production and 0.3% for $ZH$. These processes are sufficient to illustrate the issue, since $W^\mp$ and $W^\pm H$ production show very similar behavior to the $Z$ and $ZH$ processes, respectively.

Results are shown in Fig. 2. The agreement of the jettiness calculations with the normal MCFM result is excellent overall, particularly for central production $|y| \lesssim 3$. However, there is evidence for an increase in the size of the power corrections at larger absolute rapidities.\(^2\) The reason for the qualitative difference in the behavior is thus two-fold. First, the onset of power corrections with increased rapidity occurs sooner for $Z$ production. Second, and critically, the shape of the rapidity distribution is much broader for $Z$ production so that the effect of the high-rapidity tails is more apparent in the inclusive rates presented in the previous section. It suggests that a restriction to more central rapidities would decrease the effect of power corrections and speed the convergence to the correct result.

4.3 Cross sections under cuts

As an explicit demonstration of this behavior we will contrast the effect of the power corrections on the inclusive cross section with the behavior under a more realistic set of experimental cuts. Rather than cutting directly on the rapidity of the $W$ or $Z$ boson, we instead apply a minimal set of cuts on the $W$ and $Z$ boson decay products that might be applied in an experimental analysis. We consider a $Z$ boson decay to an electron–positron pair and demand that both leptons be observed in the central region, $|y(e^\pm)| < 2.5$. For the $W^+$ boson case we consider the decay into a positron and a neutrino, imposing a rapidity constraint on the charged lepton $|y(e^+)| < 2.5$ and a minimum missing transverse energy (MET) of 30 GeV. Note that the application of these cuts means that a comparison with the code ZWMS can no longer be made. Although DYNNLO [14] or FEWZ [87,88] could be used to provide a reference cross section under these cuts we do not pursue that here. Instead we simply normalize to the (fitted) asymptotic result.

The results of this study are shown in Fig. 3. As anticipated, the effect of the cuts is to significantly decrease the $T_0^\text{cut}$-dependence of the cross section. For instance, rather than a difference of approximately 1% with the asymptotic result for $T_0^\text{cut} = 0.02$ GeV in the inclusive case, the fiducial cross section differs by a few per mille or less for the same value of $T_0^\text{cut}$. The inability to restrict the rapidity of the unobserved neutrino in the case of $W^+ \rightarrow e^+ \nu$ production, $\tau_0 \sim p_T \exp(-|\eta|)$ and therefore at large rapidity the transverse momentum cutoff is larger, increasing the expected power corrections.

\(^2\) This can be expected as $\tau_0 \sim p_T \exp(-|\eta|)$ and therefore at large rapidity the transverse momentum cutoff is larger, increasing the expected power corrections.
Fig. 2 MCFM calculation using \(N\)-jettiness subtraction of the NLO contribution to the rapidity distributions of the Higgs boson (left), Z boson (center) and \(ZH\) system (right), in the \(gg \to H, Z,\) and \(ZH\) production processes, respectively. Results are shown for two values of \(T_{\text{cut}}^0\) (in GeV) and are compared with the normal MCFM calculation (solid histogram). The lower panel shows the ratio of the jettiness results to the normal MCFM calculation.

Fig. 3 The ratio of the MCFM \(N\)-jettiness calculation of the NLO coefficient to the fitted asymptotic result, as a function of the \(N\)-jettiness resolution parameter \(T_{\text{cut}}^0\) in GeV. The comparison is performed for \(Z\) (top) and \(W^+\) production (bottom) and for both the inclusive case and for a minimal set of fiducial cuts (detailed in the text). The lines represent fits to the individual points using the form given in Eq. (4.1) compared to \(Z \to e^- e^+\), leads to a slightly slower approach to the correct result.

Of course the opposite behavior is also possible. The study of some observables necessitates the application of experimental cuts that may increase the sensitivity to power corrections and thus slow the convergence to the correct result. In the context of this study it is most clear if rapidity cuts on the boson decay products are relaxed further. The choice of \(T_{\text{cut}}^0\) in the \(N\)-jettiness subtraction method should therefore be examined carefully when probing phase-space regions that may be susceptible to such effects.

5 \(N\)-Jettiness subtraction at NNLO

At NNLO, the \(N\)-jettiness subtraction method involves an above-cut contribution that corresponds to a NLO calculation of the process containing an additional parton. In contrast to the previous order, this results in genuine numerical instabilities that primarily arise from the cancellation of subtraction terms in the real radiation contribution. As a result we must use a larger value of the safety-cutoff parameter, namely \(\text{cutoff} = 10^{-8}\) GeV. This is appropriate for computations in double precision, such as the ones presented in this paper. Although we do not include any quadruple precision results here, we note that this cut may be relaxed significantly in that case. We note the caveat that the running time of the code increases significantly in quadruple precision, by about an order of magnitude.

5.1 Inclusive cross sections

The expected NNLO cross sections in the inclusive case, obtained using the already-available public codes listed previously, are shown in Table 6. The corrections to the \(gg \to H\) process are again large at this order, while all of the other processes have corrections in the 2–10% range. Of these other processes \(ZH\) production has the largest correction, but this is largely due to the effect of finite \(gg \to ZH\) and top-Yukawa contributions, as discussed previously.

The calculation of the NNLO coefficients by jettiness subtraction are compared with results from the literature in Fig. 4. Note that all of the plots use a common scale for the ordinates, which display the ratio, except for the one representing the

![Diagram](image-url)
Table 6  NNLO corrections to the processes computed in this paper

| Process | $\Delta\sigma_{\text{NNLO}}$ | $\Delta\sigma_{\text{NNLO/LO}}$ |
|---------|----------------|-------------------|
| $H$     | 10.01 pb       | 0.774             |
| $Z$     | 2.200 nb       | 0.050             |
| $W^+$   | 4.702 nb       | 0.058             |
| $ZH$    | 0.06562 pb     | 0.096             |
| $W^+H + W^-H$ | 0.0294 pb | 0.023 |

$gg \rightarrow H$ calculation, for which the power corrections are much smaller. It is clear from this figure that there is a slower approach to the asymptotic result than at NLO, but that excellent agreement is still obtained for smaller values of $T_0^{\text{cut}}$. The relatively poorer approach to the true result is expected from the behavior of the power corrections at NNLO, whose leading two terms can be modeled after integration over the final-state phase space as

$$\Delta\sigma_{\text{NNLO/\text{jettiness}}}(T_0^{\text{cut}}) = \Delta\sigma_{\text{NNLO}} + c_3 \left( \frac{T_0^{\text{cut}}}{Q} \right) \log^3 \left( \frac{T_0^{\text{cut}}}{Q} \right) + c_2 \left( \frac{T_0^{\text{cut}}}{Q} \right) \log^2 \left( \frac{T_0^{\text{cut}}}{Q} \right), \tag{5.1}$$

where $Q$ is the appropriate scale as before and $c_{2,3}$ are unknown constants. Also shown in Fig. 4 are fits of the results to Eq. (5.1), with the values of $\Delta\sigma_{\text{NNLO}}$ and $c_{2,3}$ determined in the fit. The subleading term is only important in the case of the $gg \rightarrow H$ process, in order to capture the observed turn-over for larger values of $T_0^{\text{cut}}$. For $gg \rightarrow H$, $ZH$, and $W^\pm H$ production the fit value of $\Delta\sigma_{\text{NNLO}}$ differs from the known result given in Table 6 by less than one per mille. For the $Z$ and $W^+$ processes the agreement is not as good, at the level of approximately 4%.

5.2 Rapidity distributions at NNLO

Given the effect of the power corrections on the rapidity distribution at NLO, we expect to see a similar pattern at NNLO. We compare predictions for $T_0^{\text{cut}} = 0.01$ GeV and $T_0^{\text{cut}} = 0.004$ GeV. For the $gg \rightarrow H$ and $ZH$ processes that we study here, the predictions for $T_0^{\text{cut}} = 0.004$ GeV should be a good proxy for the exact distribution given the small deviations from the inclusive cross section to which they correspond (around 0.8% for both). For $Z$ production, this value of $T_0^{\text{cut}}$ yields a total cross section that differs by 10% from the known result. To obtain an actual phenomenological result one must run with a lower $T_0^{\text{cut}}$. Nevertheless it is sufficient to demonstrate the pattern of the power corrections.

The dependence on $T_0^{\text{cut}}$ of the NNLO contributions to the rapidity distributions is illustrated in Fig. 5. As observed at NLO, all three distributions are much less sensitive to the choice of $T_0^{\text{cut}}$ in the central region than at large rapidities. The quality of the independence from $T_0^{\text{cut}}$ deteriorates substantially for $|y| \gtrsim 2$. However, even in the central region, the $Z$ process is far more affected by the choice of $T_0^{\text{cut}}$ than the other two calculations. In the more forward regions, which still contribute to the cross section at an appreciable level, the $T_0^{\text{cut}}$ dependence rises to the level of a few tens of percent. For this reason it is crucial to apply the basic fidu-
Fig. 5 MCFM calculation of the NNLO contribution to the rapidity distributions of the Higgs boson (left), Z boson (center) and ZH system (right), in the $gg \to H$, $Z$, and $ZH$ production processes, respectively. Results are shown for two values of $T^0_{\text{cut}}$, with the lower panel showing the ratio of the $T^0_{\text{cut}} = 0.01$ GeV result to the $T^0_{\text{cut}} = 0.004$ GeV one.

Fig. 6 MCFM calculation of the full NNLO result for the rapidity distributions of the Higgs boson (left), Z boson (center) and ZH system (right), in the $gg \to H$, $Z$, and $ZH$ production processes respectively. Results are shown for two values of $T^0_{\text{cut}}$, with the lower panel showing the ratio of the $T^0_{\text{cut}} = 0.01$ GeV result to the $T^0_{\text{cut}} = 0.004$ GeV one.

In order to provide a full NNLO prediction for the rapidity distributions discussed in this section we sum the results of a standard MCFM calculation at NLO and a computation of only the NNLO correction using jettiness subtraction. The resulting distributions are shown in Fig. 6. The $gg \to H$ and $Z$ production processes differ by a couple of percent in the tails of the distribution, for these two values of $T^0_{\text{cut}}$, but are otherwise in excellent agreement. The dependence on $T^0_{\text{cut}}$ is even smaller for the case of $ZH$ production.
5.3 Cross sections under cuts

Although the W and Z production cases are the most sensitive to \( T_0^{\text{cut}} \) at NNLO, at this order both \( ZH \) and \( W^{\pm}H \) production also display a non-negligible dependence on \( T_0^{\text{cut}} \). We therefore consider all four processes in this section. For \( W \) and \( Z \) production we apply the same cuts as before. For the other processes we consider the final states \( W^{\pm}(\to \ell^\pm v)H(\to \gamma\gamma) \) and \( Z(\to \ell^+\ell^-)H(\to b\bar{b}) \) but do not apply any cuts to the Higgs boson decay products in either case. In this way the results remain valid for any decay channel of the Higgs boson. The \( W^{\pm} \) and \( Z \) decay products are subject to the same cuts as in the corresponding inclusive \( W \) and \( Z \) production processes.

The results of this study are shown in Fig. 7. For the \( W \) and \( Z \) cases, the improvement is dramatic; for \( T_0^{\text{cut}} = 0.02 \) GeV the difference from the asymptotic result improves from approximately 35% in the inclusive case to 8% under cuts. A similar level of improvement applies in the case of \( W \) production. For \( ZH \) production the gain is less pronounced due to the fact that only the \( Z \) decay products are restricted in rapidity, which results in a less stringent constraint on the combined \( ZH \) system. Nevertheless, the agreement with the asymptotic result improves by about a factor of 2 relative to the case of no cuts. The asymptotic value of each NNLO \( N \)-jettiness calculation, together with the LO cross sections, is shown in Table 7.

### Table 7 NNLO corrections under the basic fiducial cuts described in the text

| Process | \( \sigma^{\text{LO},\text{fid}} \) | \( \Delta \sigma^{\text{NNLO},\text{fid}} \) | \( \Delta \sigma^{\text{NNLO},\text{fid}} / \sigma^{\text{LO},\text{fid}} \) |
|---------|----------------|----------------|----------------|
| \( Z \) | 708.6 pb | 44.8 pb | 0.063 |
| \( W^+ \) | 3.259 nb | 270 pb | 0.083 |
| \( ZH \) | 9.606 fb | 1.126 fb | 0.12 |
| \( W^{\pm}H \) | 0.1337 fb | 0.00353 fb | 0.026 |

6 Numerics

In this section we discuss the numerical performance of MCFM. As an illustration we will run the hybrid openMP/MPI version of MCFM on a modest sized cluster. This cluster consists of 24 nodes, each node having of a motherboard with two Intel X5650 chips (2.67 GHz) using an unified memory. Each of the Intel chips has 6 cores, resulting in a total of 24 \( \times \) 2 \( \times \) 6 = 288 computing cores for the cluster. The nodes are connected using InfiniBand NFS mounts.

We will use \( 4 \times 100,000 + 10 \times 1,000,000 \) VEGAS events in the remainder of this section. It is straightforward to scale the results obtained for this particular cluster to other cluster configurations. Specifically we examine two important performance issues. First, we will look at the time required to calculate the cross section as a function of the number of cores used. Second, we will look at the obtained statistical precision due to the Monte Carlo integration as a function of the \( T_0^{\text{cut}} \) parameter. For all the runs in this section we use, in addition to the input parameters of Table 2, a collision energy of 14 TeV and an inclusive anti-\( k_T \) jet algorithm with a cone size of 0.4. We apply, where applicable, the following cut on the transverse momenta of the final-state objects: \( p_T^{\text{jet}} > 20 \) GeV, \( p_T^{\ell\pm} > 25 \) GeV, \( p_T^{\text{MISS}} > 40 \) GeV, \( p_T^{JET} > 40 \) GeV and \( p_T^{\ell\pm} > 20 \) GeV. The rapidity of all final-state objects is required to be less than 2.5 and we require a separation between the observable final-state objects of \( \Delta R > 0.4 \). When a \( Z \)-boson is produced we apply the additional cut
Calculating cross sections at higher order requires a significant amount of computing power. In Ref. [57] several of us extended MCFM to use openMP by modifying VEGAS in such a manner that it distributes the event generation and evaluation over the computing cores of a single node/motherboard. By using multiple computing cores openMP makes the evaluation of NLO cross sections on desktops efficient, while still using a single VEGAS grid for the optimization of the numerical integration. For a timely evaluation of cross sections at NNLO it is desirable to use a cluster combining many processors. As the processors in a cluster do not share the same physical memory one has to use MPI. We extend VEGAS to use MPI to distribute the event generation and evaluation further over all processors, while openMP still distributes the events per processor over its computing cores. Again a single VEGAS grid is used to optimize the numerical integration. It is important to use openMP to distribute the events on a single processor as it keeps only one version of shared variables, while MPI would keep a separate copy of those variables for each MPI process thereby using the limited cache memory in an inefficient manner. This is particularly important as MCFM use large shared arrays such as for example the VEGAS grid, PDF grids, histograms, etc. which are common for all computing cores. It is therefore beneficial to maintain a hybrid openMP/MPI version of VEGAS, especially given the continuing increase of the number of cores per processor.

There are two limits which come into play when executing parallel code. The first limit is the memory bound limit. Here the evaluation time is determined by memory transfers and not by computations. In this limit the evaluation time will not scale well when adding more computing cores and improving the scaling behavior will be difficult, necessitating a better management of cache memory by the openMP code and/or more efficient message passing by the MPI code. In the other limit the evaluation time is determined by the computations and time used for memory management is negligible. In this limit the execution time will scale perfectly with the number of processors, i.e. doubling the number of processors will halve the execution time. These limits are important in order to understand the scaling behavior seen in MCFM.

The scaling of the computing time with the number of processing cores for the process \( pp \rightarrow W^+ \rightarrow l^+\nu \) is given in Table 8, with a visible representation in Fig. 8. We have the option to run one MPI process per node and let openMP distribute the events over the 12 cores of the two processors (indicated by the \( 1 \times 12 \) column). This in general is not a preferred mode of operating because the cache memory is divided over the two processors, requiring openMP to make sure the two cache memories are synchronized, unduly invoking a memory management overhead on the time needed for the evaluation. Alternatively, by running two MPI jobs per node openMP is used on a single processor thereby optimizing the cache usage and minimizing memory management.

Table 8 The time required (in seconds) to evaluate the \( pp \rightarrow W^+ \rightarrow l^+\nu \) total cross section using the given number of cores for the node openMP texture of one MPI job per processor (\( 2 \times 6 \)) and a single MPI job per node (\( 1 \times 12 \)). Results are given for LO, NLO, and NNLO total cross sections

| Cores | LO \( W^+ (2 \times 6) \) | LO \( W^+ (1 \times 12) \) | NLO \( W^+ (2 \times 6) \) | NLO \( W^+ (1 \times 12) \) | NNLO \( W^+ (2 \times 6) \) | NNLO \( W^+ (1 \times 12) \) |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| 6     | 11.15          | 308.9          | 10022          |                 |                 |                 |
| 12    | 5.99           | 8.97           | 159.1          | 171.6          |                 |                 |
| 24    | 3.45           | 5.44           | 82.2           | 89.5           |                 |                 |
| 48    | 2.25           | 3.81           | 44.8           | 48.3           |                 |                 |
| 72    | 2.01           | 3.21           | 32.1           | 36.9           |                 |                 |
| 96    | 1.96           | 2.81           | 26.4           | 33.8           |                 |                 |
| 144   | 1.95           | 2.47           | 23.7           | 29.4           |                 |                 |
| 216   | 1.93           | 2.28           | 22.9           | 26.7           |                 |                 |
| 288   | 2.20           | 2.32           | 23.3           | 25.4           |                 |                 |

Fig. 8 The evaluation time (in seconds) to calculate the total cross section for the process \( pp \rightarrow W^+ \rightarrow l^+\nu \) at LO, NLO, and NNLO as a function of the number of cores used for both the \( 1 \times 12 \) and the \( 2 \times 6 \) node texture.
The relative statistical precision (in percentages) on the Table 9. The time required (in seconds) to evaluate the $pp \rightarrow W^+ \rightarrow l^+ v$, $pp \rightarrow Z \rightarrow l^+ l^-$, $pp \rightarrow H \rightarrow \gamma \gamma$, $pp \rightarrow H + W^+ \rightarrow \gamma \gamma \gamma l^+ l^-$, and $pp \rightarrow \gamma \gamma$ cross sections at NNLO as a function of one MPI job per 6-core processor ($2 \times 6$).

| MPI jobs | $W^+$ | $Z$ | $H$ | $HW^+$ | $HZ$ | $\gamma \gamma$ |
|----------|-------|-----|-----|--------|------|--------------|
| 1        | 10022 | 20283 | 9079 | 9128 | 14357 | 27274 |
| 2        | 5068  | 10173 | 4530 | 4639 | 7222 | 13704 |
| 4        | 2559  | 5109  | 2339 | 2349 | 3655 | 6922 |
| 8        | 1326  | 2581  | 1196 | 1211 | 1846 | 3492 |
| 12       | 911   | 1752  | 821  | 823  | 1276 | 2352 |
| 16       | 702   | 1336  | 637  | 634  | 954  | 1773 |
| 24       | 585   | 915   | 466  | 452  | 662  | 1219 |
| 36       | 381   | 643   | 360  | 340  | 473  | 839 |
| 48       | 328   | 505   | 323  | 283  | 378  | 657 |

Fig. 9 The evaluation time (in seconds) needed to calculate the total cross section for the processes $pp \rightarrow W^+ \rightarrow l^+ v$, $pp \rightarrow Z \rightarrow l^+ l^-$, $pp \rightarrow H \rightarrow \gamma \gamma$, $pp \rightarrow H + W^+ \rightarrow \gamma \gamma l^+ l^-$, and $pp \rightarrow \gamma \gamma$ at NNLO as a function of the number of MPI processes used (each MPI process uses openMP on 6 cores)

Table 10 The relative statistical precision (in percentages) on the $pp \rightarrow W^+ \rightarrow l^+ v$, $pp \rightarrow Z \rightarrow l^+ l^-$, $pp \rightarrow H \rightarrow \gamma \gamma$, $pp \rightarrow H + W^+ \rightarrow \gamma \gamma l^+ l^-$, and $pp \rightarrow \gamma \gamma$ cross sections at NNLO as a function of $T_0^\text{cut}$ (in GeV) using $4 \times 2 \times 6$ cores. Also given in brackets is the evaluation time (in seconds).

| $T_0^\text{cut}$ | $W^+$ | $Z$ | $H$ | $HW^+$ | $HZ$ | $\gamma \gamma$ |
|------------------|-------|-----|-----|--------|------|--------------|
| 0.001            | 2% (1397) | 0.9% (2770) | 0.05% (1256) | 10% (1263) | 6% (1939) | 0.4% (3706) |
| 0.005            | 0.7% (1358) | 0.4% (2701) | 0.04% (1234) | 3% (1238) | 2% (1906) | 0.2% (3661) |
| 0.01             | 0.5% (1356) | 0.2% (2677) | 0.04% (1214) | 2% (1222) | 1% (1847) | 0.15% (3585) |
| 0.05             | 0.2% (1315) | 0.08% (2572) | 0.04% (1197) | 0.6% (1206) | 0.4% (1841) | 0.09% (3492) |
| 0.1              | 0.09% (1307) | 0.05% (2526) | 0.04% (1186) | 0.3% (1186) | 0.2% (1847) | 0.08% (3427) |
| 0.5              | 0.04% (1266) | 0.04% (2356) | 0.04% (1176) | 0.1% (1150) | 0.09% (1768) | 0.07% (3376) |

Table 11 Values of $T_0^\text{cut}$ (in GeV) required to perform the NNLO $N$-jetiness calculation to a given accuracy, for the processes studied in this paper. At larger $T_0^\text{cut}$ the accuracy deteriorates because of increased power corrections.

| Process | $\sigma^{NLO}$ | 1% accuracy | 0.2% accuracy |
|---------|----------------|-------------|---------------|
| $gg \rightarrow H$ | Inclusive | 0.03 | 0.002 |
| $Z$ | Inclusive | 0.01 | 0.002 |
| $W^+$ | Inclusive | 0.005 | 0.001 |
| $ZH$ | Inclusive | 0.3 | 0.02 |
| $W^\pm H$ | Inclusive | 0.2 | 0.01 |
| $\gamma \gamma$ | Cuts [33] | 0.01 | 0.001 |

overhead ($2 \times 6$ column). This is clearly demonstrated in Table 8. For example using all 288 processors on the cluster we can use of 48 MPI jobs (2 MPI jobs per node) to evaluate the NNLO cross section in 328 seconds, or use 24 MPI jobs (1 MPI job per node) requiring 358 seconds to evaluate the NNLO cross section. The time difference is due to the fractured cache memory caused by forcing openMP to use two processors in the case of running with 24 MPI jobs. Therefore in the remainder of this section we will use 2 MPI jobs per node, allowing openMP to operate on a single processor.

Especially at LO, and to some extent at NLO, the computation effort to evaluate this process is minimal, making the execution time operate close to the memory bound limit. This behavior is exhibited in Table 8 and Fig. 8 where for LO the evaluation time no longer improves when using more than 12 MPI jobs (and for NLO more than 18 MPI jobs). At NNLO using more cores still improves the evaluation time, as a consequence of the need to evaluate a large number of the more computational intensive double parton bremsstrahlung events. It is worth noting that the execution time on a single processor using openMP executes in under 3 h making the evaluation of this process on desktops very feasible. Running on the full cluster using the 48 processors results in an execution time of less than 4 min. This means one can easily
Fig. 10 The NLO $T_0^{cut}$ dependence for $gg \rightarrow H$ of Fig. 1 is shown as the red line in the upper pane. The purple line gives the remaining $T_0^{cut}$ dependence when the analytically calculated expression for the leading power correction is added to MCFM. The lower pane gives the same at NNLO for $gg \rightarrow H$ where the red line is taken from Fig. 4.

increase the number of events and lower the tau cut value to obtain better statistics.

Next we look at all the new NNLO processes added to MCFM in Table 9 and Fig. 9 where the time in seconds is given as a function of the number of MPI jobs (= number of processors) used. As can be seen, the processes scale well all the way up to the 288 processors. Some indication of a less than perfect scaling can be seen in the simplest of the NNLO processes $pp \rightarrow H \rightarrow \gamma \gamma$ when we get to a high number of processors indicating there is some memory overhead. All other processors still are computing dominated, which will allow easy speed up by invoking even more processors. The most complicated NNLO process $pp \rightarrow \gamma \gamma$ takes just under 11 min to evaluate using 48 processors. Therefore obtaining higher statistics is rather easy. This process would still only take a bit less than 8 h on a single processor desktop.

Finally, the statistical integration error obtained for the inclusive cross section given the cuts using the 10,000,000 VEGAS events as a function of the $T_0^{cut}$ is given in Table 10. The evaluation times are given for using 8 processors. As can be seen there is a small dependence of the evaluation time on the choice of the $T_0^{cut}$. As we choose the $T_0$ cut smaller the Monte Carlo becomes more “efficient” because it will generate more soft/collinear events. That is, less events will be rejected by the cuts hence the evaluation time will grow.

As can be seen from Table 10, the acquired statistical uncertainty is quite process dependent. However, the value of the $T_0^{cut}$ will also determine the systematic error due to the power corrections. Looking at Table 11 we see the required value of $T_0^{cut}$ to reduce the power corrections to a 1% or a 0.2% level. First, focusing on the 1% uncertainty we see that in all cases statistical error obtained with the 10,000,000 events is smaller than 1%. The worst case is the inclusive $W^+$ production with a statistical uncertainty of 0.7%. For all other cases the statistical error is more than on order of magnitude smaller. To achieve a systematic error of about 0.2% we see that we need to reduce the statistical uncertainty significantly in order to be smaller than the systematic error. The reduction for some processes is about an order of magnitude, requiring of the order of 100 times more events. This means that an overall uncertainty of order 1% is easily obtainable using a desktop, however, going to the per-mille level will require a modest computer cluster such as the one used for the numerical results in this section.

3 The size of the power corrections for the diphoton process is obtained from the results of Ref. [33].
7 Summary

In this paper we presented details of the performance of the first NNLO version of MCFM. Using the non-local N-jettiness subtraction method, we included the NNLO corrections for six final states: $pp \to W^\pm$, $pp \to Z/\gamma^*$, $pp \to H$, $pp \to W^\pm H$, $pp \to ZH$ and $pp \to \gamma\gamma$. For each process decays of the unstable vector bosons are included where appropriate. The method was checked at NLO against existing calculations and excellent agreement was found. At NNLO the dependence on the jettiness cut was studied in great detail and some guidelines on the choice of the jettiness cut were given for all NNLO processes added to MCFM.

Another addition to MCFM is the ability to run in a hybrid openMP/MPI mode, enabling the Monte Carlo to use clusters efficiently while still maintaining a single VEGAS grid. The evaluation time for NNLO inclusive cross sections with an overall precision better than 1% on a single 8-core processor using openMP ranges from 3–8 h depending on the specific process. It was shown all processes scaled well on a multi-processor cluster using MPI in addition to openMP, giving an evaluation time of 5–11 min on a 48 8-core processor cluster. A small cluster of 50+ cores will give good statistics for distributions at NNLO in a short time-frame.

The N-jettiness subtraction method in MCFM is now well tested. Combined with the hybrid openMP/MPI option, MCFM is now ready to include in the near future more complicated processes including jet final states. Going forward, it would be highly beneficial if one could calculate some of the power corrections analytically in order to operate at higher values of $T_0$ and thus speed up the code. Results obtained from a preliminary calculation of the dominant power correction for the total $gg \to H$ cross section at NLO and NNLO are shown in Fig. 10. This calculation has been performed as a function of the rapidity of the Higgs boson, the details of which will be presented (including also other Drell–Yan like processes) in a separate publication [58]. As can be seen from Fig. 10 the effect of including the dominant power correction term in MCFM is substantial and one can choose $T_0$ an order of magnitude larger, while still retaining the same distance from the asymptotic result. This could have a large impact on the ultimate achievable precision of the jettiness method implemented in MCFM.

Acknowledgements RKE would like to thank Robert Harlander for helpful discussions. R. B. is supported by the DOE contract DE-AC02-06CH11357. J. C. and W. G. are supported by the DOE contract DE-AC02-07CH11359. C. F. is supported by the NSF Grant PHY-1520916. X. L. is supported by the DOE Grant DE-FG02-93ER-40762. F. P. is supported by the DOE Grants DE-FG02-91ER40684 and DE-AC02-06CH11357. Support was provided by the Center for Computational Research at the University at Buffalo. R. B., W. G., X. L., and F. P. thank the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara, which is supported by the U. S. National Science Foundation under Grant No. NSF PHY11-25915. C. W. is supported U. S. National Science Foundation under Grant No. NSF PHY-1619877

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP3.

References

1. C. Anastasiou, K. Melnikov, F. Petriello, Higgs boson production at hadron colliders: Differential cross sections through next-to-next-to-leading order. Phys. Rev. Lett. 93, 262002 (2004). arXiv:hep-ph/0409088
2. C. Anastasiou, K. Melnikov, F. Petriello, Fully differential Higgs boson production and the di-photon signal through next-to-next-to-leading order. Nucl. Phys. B 724, 197–246 (2005). arXiv:hep-ph/0501130
3. S. Catani, M. Grazzini, An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. Phys. Rev. Lett. 98, 222002 (2007). arXiv:hep-ph/0703012
4. J. Gaunt, M. Stahlhoven, F.J. Tackmann, J.R. Walsh, N-jettiness Subtractions for NNLO QCD Calculations. JHEP 09, 058 (2015). arXiv:1505.04794
5. R. Bouguhezal, F. Caola, K. Melnikov, F. Petriello, M. Schulze, Higgs boson production in association with a jet at next-to-next-to-leading order in perturbative QCD. JHEP 06, 072 (2013). arXiv:1302.6216
6. X. Chen, T. Gehrmann, E.W.N. Glover, M. Jaquier, Precise QCD predictions for the production of Higgs + jet final states. Phys. Lett. B 740, 147–150 (2015). arXiv:1408.5325
7. R. Bouguhezal, F. Caola, K. Melnikov, F. Petriello, M. Schulze, Higgs boson production in association with a jet at next-to-next-to-leading order, Phys. Rev. Lett. 115(8), 082003 (2015). arXiv:1504.07922
8. R. Bouguhezal, C. Focke, W. Giele, X. Liu, F. Petriello, Higgs boson production in association with a jet at NNLO using jettiness subtraction, Phys. Lett. B748, 5–8 (2015). arXiv:1505.03893
9. X. Chen, T. Gehrmann, N. Glover, M. Jaquier, Higgs plus one jet production at NNLO, 2016. arXiv:1604.04085
10. M. Cacciari, F. A. Dreyer, A. Karlberg, G. P. Salam, G. Zanderighi, Fully differential vector-Boson-Fusion Higgs production at next-to-next-to-leading order, Phys. Rev. Lett. 115(8), 082002 (2015). arXiv:1506.02660
11. C. Anastasiou, F. Herzog, A. Lazopoulos, The fully differential decay rate of a Higgs boson to bottom-quarks at NNLO in QCD. JHEP 03, 035 (2012). arXiv:1110.2368
12. V. Del Duca, C. Duhr, G. Somogyi, F. Tramontano, Z. Trócsányi, Higgs boson decay into b-quarks at NNLO accuracy. JHEP 04, 036 (2015). arXiv:1501.07226
13. K. Melnikov, F. Petriello, The W boson production cross section at the LHC through $O(\alpha_s^7)$, Phys. Rev. Lett. 96, 231803 (2006). arXiv:hep-ph/0603182
14. S. Catani, L. Cieri, G. Ferrera, D. de Florian, M. Grazzini, Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO. Phys. Rev. Lett. 103, 082001 (2009). arXiv:0903.2120
15. K. Melnikov, F. Petriello, Electroweak gauge boson production at hadron colliders through $O(\alpha(s)^{n+2})$, Phys. Rev. D 74, 114017 (2006). hep-ph/0609070
60. M. Beneke, A.P. Chapovsky, M. Diehl, T. Feldmann, Soft collinear effective theory and heavy to light currents beyond leading power. Nucl. Phys. B 643, 431–476 (2002). arXiv:hep-ph/0206152

61. C.W. Bauer, S. Fleming, M.E. Luke, Summing Sudakov logarithms in B –> X(s gamma) in effective field theory. Phys. Rev. D 64, 014006 (2001). arXiv:hep-ph/0005275

62. C.W. Bauer, S. Fleming, D. Pirjol, I.W. Stewart, An effective field theory for collinear and soft gluons: heavy to light decays. Phys. Rev. D 63, 114020 (2001). arXiv:hep-ph/0011336

63. C.W. Bauer, I.W. Stewart, Invariant operators in collinear effective theory. Phys. Lett. B 516, 134–142 (2001). arXiv:hep-ph/0107001

64. C.W. Bauer, D. Pirjol, I.W. Stewart, Soft collinear factorization in effective field theory. Phys. Rev. D 65, 054022 (2002). arXiv:hep-ph/0109045

65. C.W. Bauer, S. Fleming, D. Pirjol, I.Z. Rothstein, I.W. Stewart, Hard scattering factorization from effective field theory. Phys. Rev. D 66, 014017 (2002). arXiv:hep-ph/0202088

66. T. Becher, M. Neubert, Toward a NNLO calculation of the anti-B –> X(s) gamma decay rate with a cut on photon energy. II. Two-loop result for the jet function. Phys. Lett. B 637, 251–259 (2006). arXiv:hep-ph/0605367

67. J.R. Gaunt, M. Stahlhofen, F.J. Tackmann, The quark beam function at two loops. JHEP 04, 113 (2014). arXiv:1401.5478

68. R. Gavin, Y. Li, F. Petriello, Combining QCD and electroweak corrections to dilepton production in FEWZ. Phys. Rev. D 86, 094034 (2012). arXiv:1208.5967