HIGGS BOSON PAIR PRODUCTION: MONTE CARLO GENERATOR INTERFACE AND PARTON SHOWER*

S.P. Jones

Max Planck Institute for Physics, Föhringer Ring 6, 80805 München, Germany

(Received February 5, 2018)

We discuss the combination of the NLO QCD matrix element for double Higgs boson production with a parton shower, including the full top-quark mass dependence. Since the 2-loop double Higgs boson production amplitude is currently known only numerically, in order to produce a fast and stable code for the evaluation of the virtual matrix element, we construct a grid based on a fixed number of precomputed phase-space points. Results are generated in both the POWHEG-BOX and MadGraph5_aMC@NLO Monte Carlo frameworks and showered with Pythia. We investigate the sensitivity of showered predictions of the Higgs boson pair transverse momentum distribution to the parameters of the POWHEG matching scheme.

DOI:10.5506/APhysPolBSupp.11.295

1. Introduction

The exploration of the Higgs sector is one of the major goals for the next phases of the LHC experiments. In particular, the form of the Higgs potential as predicted by the Standard Model (SM) remains to be confirmed. In the SM, after electroweak symmetry breaking, the Higgs boson self-interactions are predicted in terms of the Higgs boson mass and vacuum expectation value which are known precisely. However, the Higgs boson self-coupling is still only very weakly constrained experimentally; its precise measurement will, therefore, serve as a new and important test of the SM. The self-coupling can be measured, for example, via Higgs boson pair production in gluon fusion, which is the dominant production mechanism of Higgs boson pairs. However, the cross section is around 1000 times smaller than that for single Higgs production, which makes the measurement very challenging even with the high luminosity upgrade of the LHC. Nevertheless, delicate cancellations between different contributions which happen in the

SM are altered in most new physics models, leading to potentially large effects, which makes this channel very interesting for present day new physics searches.

Theoretically, double Higgs boson production in gluon fusion is challenging to compute in part due to the fact that it is a loop-induced process with several mass scales. The leading order (LO) 1-loop result was first computed some time ago [1, 2]. Beyond LO, until recently, results were known only in the Higgs Effective Theory (HEFT). In the HEFT, computations are performed in the infinite top-quark mass limit, which induces effective (multi-)gluon-Higgs couplings. This, in turn, produces a tree-level LO and reduces the number of scales appearing in the calculation. In the HEFT, the next-to-leading order (NLO) 1-loop results are known and have been reweighted by the Born result in the SM ("Born improved"), this procedure includes the effect of the top-quark mass at LO [3, 4]. The NLO $K$-factor, defined as the ratio of the NLO result to the LO result, in the HEFT is approximately 2. The next-to-next-to-leading order (NNLO) 2-loop results in the HEFT are also known [5–8] and have recently been computed also differentially [9]. The NNLO cross section is approximately 20% larger than at NLO. Threshold resummation has also been applied to the HEFT results at NLO+NNLL [10] and NNLO+NNLL [11] accuracy. Additionally, an expansion in the top-quark mass is known up to $1/m^2_t$ at NLO and $1/m^4_t$ at NNLO [8, 12]. An alternative approximation is to compute the real radiation in the SM but use the NLO HEFT virtual matrix element rescaled event-by-event by the Born result in the SM, this is known in the literature as "Full Theory (FT) Approx" [13, 14]. In the SM, the NLO (2-loop) result including the full top-quark mass is known only numerically [15–18]. For $\sqrt{s} = 14$ TeV, the SM cross section is 14% below the Born Improved HEFT result and 4% below the "FT Approx" result. NLO results including an NLL resummation for the transverse momentum of the Higgs boson pair are also known including the full top-quark mass [19].

In these proceedings, based on the work presented in Ref. [20], we examine the impact of applying a parton shower (PS) to the NLO QCD result for double Higgs boson production, including the full top-quark mass dependence. In Sec. 2, we describe the setup of the Monte Carlo frameworks and the interface between the public codes and the fixed order result. We present our phenomenological predictions for NLO+PS in Sec. 3. Finally, in Sec. 4, we discuss the impact of the POWHEG $h_{\text{damp}}$ parameter on our predictions.

2. Setup

The most complicated piece of the NLO QCD contribution to double Higgs boson production is the 2-loop virtual matrix element. In this work, for the fixed order result, we rely on our earlier calculation described in detail
in Refs. [15–18]. The amplitude there was obtained both in Reduze [21] and in an extended version of GoSam capable of handling multi-loop processes. The (partial) integral reduction was performed using Reduze and the resulting integrals were evaluated numerically using SecDec [22]. In Fig. 1, we show example diagrams contributing to the 2-loop amplitude.

![Example 2-loop diagrams](image)  
Fig. 1. Example 2-loop diagrams entering the $gg \rightarrow HH$ virtual amplitude.

For the computation of the 1-loop real radiation, our POWHEG code relies on GoSam [23] which utilises QGRAF [24], FORM [25] and Spinney [26] for the evaluation of the Feynman diagrams, and offers a choice of Samurai [27, 28], golem95C [29–31] and Ninja [32,33] for the tensor reduction. At run time, the amplitudes were computed using Ninja with OneLOop [34] for the evaluation of the scalar 1-loop integrals. Within MG5_aMC@NLO, the 1-loop Born and real amplitudes are computed using MadLoop [35], which relies on CutTools [36], Ninja and Collier [37], together with an in-house implementation of the OpenLoops optimisation [38].

The fixed order NLO double Higgs result utilises numerical integration for the calculation of the master integrals appearing in the amplitude, this causes it to be computationally expensive to evaluate. The amplitude depends on two form factors, $F_1$ and $F_2$; setting an accuracy goal of 3% for the dominant form factor $F_1$ and a goal of 5–20% for $F_2$ (depending on the ratio $F_2/F_1$), the evaluation of the amplitude takes between 80 minutes and 2 days (median 2 hours) of Graphics Processing Unit (GPU) wall-clock time per phase-space point. However, the virtual matrix element depends only on the two Mandelstam invariants $\hat{s}$, $\hat{t}$ (along with the masses $m_T$, $m_H$, which are fixed to 173 GeV and 125 GeV, respectively, in the fixed order calculation). Rather than interface the amplitude directly from the Monte Carlo generators, we instead chose to build a two-dimensional grid of pre-calculated phase-space points and define an interpolation procedure, this greatly reduces the time required to evaluate the amplitude by the Monte Carlo generators. For this work, we utilise a grid based on 3741 phase-space points.

To construct the grid, we first transform the Mandelstam invariants $\hat{s}$ and $\hat{t}$ to new variables $(x, c_\theta)$ according to:

1 Measured using an Nvidia Tesla K20X.
\[ x = f(\beta(\hat{s})), \quad \text{with} \quad \beta = \left(1 - \frac{4m_h^2}{\hat{s}}\right)^{\frac{1}{2}}, \quad (1) \]

\[ c_\theta = |\cos \theta| = \left| \frac{\hat{s} + 2\hat{t} - 2m_h^2}{\hat{s}\beta(\hat{s})} \right|. \quad (2) \]

We choose \( f(\beta) \) according to the cumulative distribution function of the phase-space points used in our original calculation. This procedure results in a nearly uniform distribution of points in the \((x, c_\theta)\) unit square. Next, we apply a two-step interpolation procedure:

- Choose equidistant grid points in the \((x, c_\theta)\) unit square and estimate the result at each grid point using linear interpolation of the amplitude results in the vicinity;
- Apply Clough-Tocher interpolation [39] (as implemented in \texttt{scipy} [40]) to the results at the equidistant grid points to estimate the amplitude at any arbitrary sampling point.

This procedure reduces the size of interpolation artefacts which appear due to the numerical uncertainty of our 2-loop input data points. The python grid is publicly available and can be interfaced from \texttt{FORTRAN}, C and C++ via the python/C API [41].

**3. Results**

We now present phenomenological predictions produced by applying a parton shower to the NLO results retaining the full top-quark mass dependence. Results are produced using both the \texttt{POWHEG} [42] matching scheme, as implemented in \texttt{POWHEG-BOX} [42,43], and the \texttt{MC@NLO} [44] matching scheme, as implemented in \texttt{MadGraph5_aMC@NLO} [45,46]. All results are computed using the \texttt{PDF4LHC15_nlo_30_pdfas} [47] parton distribution function (PDF) set, along with the corresponding value for \( \alpha_s \), accessed via the \texttt{LHAPDF} [48] interface. We consider only on-shell Higgs boson production and do not include the decay of the Higgs bosons. Jets are required to have transverse momentum greater than \( p_{\text{jet}}^{\text{T,min}} = 20 \text{ GeV} \), they are clustered with the anti-\(k_T\) algorithm [49] as implemented in the \texttt{Fastjet} package [50], with jet radius \( R = 0.4 \). The scale uncertainty is estimated via a 7-point scale variation of the factorization scale \( \mu_F \) and the renormalization scale \( \mu_R \) around the central scale choice \( \mu_0 = m_{hh}/2 \). The PDF uncertainty is estimated using the error PDFs contained in the \texttt{PDF4LHC15_nlo_30_pdfas} set, it is well below the scale variation uncertainty and never exceeds 6%. The uncertainty bands shown on our results originate from scale variations only. For all our showered results, produced using either \texttt{POWHEG} or \texttt{MadGraph5_aMC@NLO}, we generate showered events using \texttt{Pythia 8.2} [51,52] with the same settings.
We begin by comparing our fixed order NLO results to our default POWHEG results produced setting $h_{\text{damp}} = 250$ GeV (in Sec. 4, we will explore the impact of this choice). In Fig. 2, we show the Higgs boson pair invariant mass distribution and the transverse momentum distribution of a randomly chosen Higgs boson for both the fixed order and the showered calculation, for these variables our fixed order result is NLO accurate. The invariant mass distribution is rather insensitive to the parton shower. The shower moderately enhances the tail of the (randomly chosen) Higgs boson $p_T$ distribution, but remains close to the upper edge of the scale variation up to $p_T^h \sim 600$ GeV.

In Fig. 3, we display the transverse momentum distribution of the leading jet and the Higgs boson pair. In the NLO fixed order $gg \rightarrow HH$ result for $p_{T1} > 0$ GeV or $p_{Thh} > 0$ GeV these distributions are populated by the real radiation and are therefore only LO accurate. Furthermore, since the jet recoils only against the Higgs boson pair, the two fixed order distributions are identical. However, the parton shower generates additional jets which decorrelate these distributions. The parton shower has a sizeable effect on these observables. In particular, the parton shower differs significantly from the fixed order result in the Sudakov region (low $p_{T1}$ or $p_{Thh}$), where the fixed order result is not reliable and a parton shower (or another technique, such as analytic resummation) must be applied to produce physically reasonable results. At large $p_{Thh}$, we see that the distribution produced using the POWHEG matching scheme plus parton shower is significantly enhanced compared to the fixed order prediction. In contrast, for large values of $p_{Thh}$, the MadGraph5_aMC@NLO result, produced using the MC@NLO matching scheme, agrees with the fixed order result. Here, the MadGraph5_aMC@NLO shower scale is chosen randomly in the interval $[0.1H_T/2, H_T/2]$, where $H_T$ is the...
scalar sum of the transverse masses of the outgoing particles computed with the Born kinematics. The result is sensitive to the choice of the shower starting scale, for example, choosing the shower scale randomly in the interval $[0.1 \sqrt{s}, \sqrt{s}]$ gives results consistent with the POWHEG results.

Fig. 3. Comparison of fixed order result at $\sqrt{s} = 14$ TeV with showered results. (Left panel) Leading jet transverse momentum distribution $p_{T}^{j_{1}}$ compared with showered results from POWHEG-BOX. (Right panel) Higgs boson pair transverse momentum distribution $p_{T}^{hh}$ compared with showered results from POWHEG-BOX and MadGraph5_aMC@NLO.

The POWHEG and MCONLO matching schemes, when coupled with a parton shower, are constructed to allow for the radiation of additional (typically) low energy/low $p_{T}$ jets and to reliably describe the (low $p_{T}$) Sudakov region without spoiling the fixed order accuracy of the observables. The matching schemes, however, can introduce differences that are formally higher order in the perturbative expansion parameter (here $\alpha_{s}$). In Sec. 4, we will tentatively motivate that the differences between the fixed order and parton shower results in Fig. 3 are similar in size to the higher order corrections and discuss our choice of the $h_{\text{damp}}$ parameter.

### 4. $h_{\text{damp}}$ and matching

We now examine results at the Les Houches event (LHE) level, i.e. after the first hard emission is weighted with the Sudakov factor according to the POWHEG method but before a parton shower is applied. Note that in order to produce physical predictions, the LHE level events must first be showered, we examine these events here only to disentangle the impact of the shower from that due to the POWHEG matching.

Within POWHEG-BOX the amount of hard radiation which is exponentiated in the Sudakov factor can be limited by changing the $h_{\text{damp}}$ parameter. This parameter allows to divide the contributions of the real radiation $R$ which are exponentiated into a singular part $R_{\text{sing}}$ and a regular part $R_{\text{reg}}$ as follows:
\[ R_{\text{sing}} = R \times F, \quad R_{\text{reg}} = R \times (1 - F), \]  

\[ F = \frac{h_{\text{damp}}^2}{(p_{T hh}^h)^2 + h_{\text{damp}}^2}. \]

The choice \( h_{\text{damp}} = \infty \text{ GeV} \) implies \( F = 1 \) and that all real radiation is exponentiated.

In Fig. 4, we compare LHE level results produced by setting, \( h_{\text{damp}} = \infty \text{ GeV} \), with predictions produced using \( h_{\text{damp}} = 250 \text{ GeV} \). The left plot shows predictions in the HEFT, whereas the right plot shows results in the full Standard Model. We observe that in both cases above 500 GeV, the LHE curve with \( h_{\text{damp}} = \infty \text{ GeV} \) significantly overshoots the NLO result. Choosing the lower value of \( h_{\text{damp}} = 250 \text{ GeV} \) suppresses the enhancement for larger values of \( p_{T hh}^h \) such that the LHE result reproduces the fixed order prediction.

One is led to wonder why the matching scheme (especially with \( h_{\text{damp}} = \infty \text{ GeV} \)) has such a large impact on the high \( p_{T hh}^h \) distribution for double Higgs production. In fact, analogous effects have already been observed in several other similar processes including \( gg \to H \) [53], single-top production [54], \( gg \to ZZ \) [55] and \( gg \to HZ \) [56]. As pointed out in Ref. [53], the \( gg \to H \) process has a large \( K \)-factor, thus higher order effects (in \( \alpha_s \)) have a considerable numerical impact on the prediction, this is also true for the other processes listed above and for double Higgs production.

To display the impact of higher order corrections, the authors of Ref. [53] compared their showered NLO result to the fixed order NNLO result and found reasonable agreement, an indication that indeed the matching/parton
shower uncertainties were similar in size to the higher order corrections. Furthermore, at large $p_T$, where the Sudakov form factor approaches 1, the POWHEG differential cross section for single Higgs production can be written as

\[ d\sigma = \left[ \tilde{B}(\bar{\Psi}_1) \frac{R(\bar{\Psi}_2)}{B(\bar{\Psi}_1)} + \sum_q R_{q\bar{q}}(\bar{\Psi}_2) \right] d\bar{\Psi}_1 d\Psi_{\text{rad}}, \]  

(5)

where $R$ represents the real corrections and $B$ is the Born contribution and $\tilde{B}$ is the Born plus NLO correction, in the gluon–gluon and gluon–quark channels. Here, the symbols $\bar{\Psi}_1$ and $\bar{\Psi}_2$ represent the Born and real phase spaces respectively, whilst $\Psi_{\text{rad}}$ parametrises the additional phase space available to the real radiation. Since $\tilde{B}/B = 1 + O(\alpha_s)$, replacing this ratio by 1 amounts to changing the cross section by higher order terms. It was found in Ref. [53] that such a replacement causes the POWHEG prediction to again agree with the NLO result at large $p_T$, demonstrating that the discrepancy comes from the matching procedure.

Very recently, a detailed study of parton shower effects was performed for double Higgs boson production using the MC@NLO matching scheme [57] as implemented in Sherpa [58]. A similar $\tilde{B} \rightarrow B$ replacement (which has a somewhat different meaning in the MC@NLO matching scheme) was shown to reproduce the NLO fixed order result for the $p_{hhT}^h$ distribution. Replacing $\tilde{B}$ with $B$ does, however, spoil the NLO accuracy of inclusive observables such as $m_{hh}$ and $p_{T}^h$.

To further examine this effect in the context of double Higgs production, it is interesting to compare our NLO+PS predictions obtained with POWHEG to the full NNLO predictions, where they are available. Indeed if we restrict ourselves to results in the HEFT the differential NNLO results

![Fig. 5. Comparison of the NNLO results from Ref. [9] with POWHEG LHE level predictions produced with (left panel) $h_{\text{damp}} = \infty$ GeV and (right panel) $h_{\text{damp}} = 250$ GeV for the Higgs-pair transverse momentum $p_{T}^{hh}$.](image-url)
are known [9]. In Fig. 5, we display the comparison between the NLO, NNLO and (left panel) $h_{\text{damp}} = \infty$ GeV LHE events and (right panel) $h_{\text{damp}} = 250$ GeV LHE events in the HEFT. We can see that indeed the $h_{\text{damp}} = \infty$ GeV LHE HEFT result is similar to the NNLO HEFT result, while the $h_{\text{damp}} = 250$ GeV result reproduces the NLO result for large $p_{T}^{hh}$.

We should point out, however, that these considerations, made using the HEFT approximation, may not carry over to the full calculation (where NNLO predictions are not available). For example, it is well-known that the HEFT approximation does not have the correct scaling behaviour at large transverse momenta.

5. Conclusion

We have presented the combination of the full NLO prediction for Higgs boson pair production, including the top-quark mass dependence at 2-loops, with a parton shower. Results are generated using two frameworks, POWHEG-BOX and MG5_aMC@NLO, in each case applying a Pythia 8.2 shower. Individual phase-space points of the 2-loop amplitude, which depends only on two independent kinematic invariants once the top-quark and Higgs boson masses are fixed, have been used to create a grid and combined with an interpolation framework, such that a value for the amplitude can be obtained at any phase-space point without re-evaluating the loop integrals.

The impact of the parton shower on the transverse momentum distribution of one Higgs boson, $p_{T}^{h}$, is moderate and the features of the various approximations that have appeared previously in the literature are preserved by the shower. The impact of the shower on the transverse momentum distribution of the Higgs boson pair, $p_{T}^{hh}$, is fairly large and differs depending on the matching scheme used, with the showered results produced using the POWHEG matching scheme remaining above the fixed order prediction even for $p_{T}^{hh} > 400$ GeV. We discuss some of the reasons for this behaviour, in particular, the fact that the tail is predicted at the first non-trivial order in the fixed order calculation and that the higher order perturbative effects are likely to be large. The dependence of the result on the $h_{\text{damp}}$ parameter of the POWHEG matching scheme is examined and we have shown how for smaller values of $h_{\text{damp}}$, the LHE level results reproduce the fixed order results for large $p_{T}^{hh}$.

Nevertheless, we observe that the inclusion of the full mass dependence, in general, has a more significant impact on the distributions relevant to Higgs boson pair production than effects coming from different shower matching schemes or shower starting scales, especially for NLO accurate observables.

---

\[^{2}\] We are grateful to Javier Mazzitelli for providing us the NNLO predictions shown in the comparisons of this section.
I would like to thank G. Heinrich, M. Kerner, G. Luisoni and E. Vryonidou for their collaboration on this project. I would also like to thank S. Borowka, N. Greiner, G. Heinrich, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke for their contribution to the double Higgs fixed order NLO project. This work was supported in part by the Research Executive Agency (REA) of the European Union under the Grant Agreement PITN-GA2012316704 (HiggsTools). The support and resources provided by the Max Planck Computing and Data Facility (MPCDF) are gratefully acknowledged.

REFERENCES


