Measurements of $t\bar{t}$ in association with charm quarks at 13 TeV with the ATLAS experiment

Knut Zoch[1*], on behalf of the ATLAS Collaboration

1 Laboratory for Particle Physics and Cosmology, Harvard University,
Cambridge, Massachusetts 02138, USA

⋆ kzoch@g.harvard.edu

*The 17th International Workshop on
Top Quark Physics (TOP2024)
Saint-Malo, France, 22-27 September 2024*

Abstract

This talk presents the ATLAS Collaboration's first measurement of the inclusive cross-section for top-quark pair production in association with charm quarks. Using the full Run 2 proton–proton collision data sample at $\sqrt{s} = 13$ TeV, collected with the ATLAS experiment at the LHC between 2015 and 2018, the measurement selects $t\bar{t}$ events with one or two charged leptons and at least one additional jet in the final state. A custom flavour-tagging algorithm is employed to simultaneously identify b-jets and c-jets. The fiducial cross-sections for $t\bar{t}+\geq 2c$ and $t\bar{t}+1c$ production are found to largely agree with predictions from various $t\bar{t}$ simulations, though all underpredict the observed values.

1 Introduction

Top-quark pair production ($t\bar{t}$) with additional heavy-flavour jets is a large irreducible background to many other rare processes predicted by the Standard Model (SM) of particle physics. Prominent examples include $t\bar{t}H$ production with $H \rightarrow b\bar{b}$ decays and the production of four top quarks ($t\bar{t}t\bar{t}$) in single-lepton or dilepton final states. In this context, *heavy flavour* refers to jets originating from b-quarks or c-quarks, labelled as b-jets and c-jets, respectively. In $t\bar{t}$ events, additional heavy-flavour jets can arise from gluon splitting into $b\bar{b}$ and $c\bar{c}$ pairs. The $b\bar{b}$ or $c\bar{c}$ pair can form separate jets ($t\bar{t}+b\bar{b}/c\bar{c}$) or be clustered into a single jet ($t\bar{t}+1B/C$). In addition, a single additional b-quark or c-quark can originate from the initial state ($t\bar{t}+1b/t\bar{t}+1c$). Illustrative Feynman diagrams for these processes are shown in Figure 1. Computations of $t\bar{t} + b\bar{b}$ production cross-sections exist at next-to-leading order (NLO) accuracy in quantum chromodynamics (QCD), but uncertainties in the choice of the renormalisation and factorisation scales remain sizeable due to the different energy scales involved in the process. Currently, no dedicated $t\bar{t} + c\bar{c}$ computations are available, emphasising the need for experimental measurements to improve the understanding of these final states.
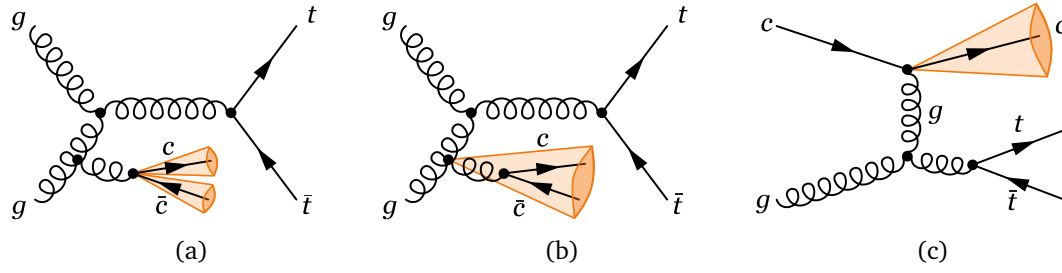
Figure 1: Illustrative Feynman diagrams for $t\bar{t}$ pair production with one or multiple additional c-jets: (a) $t\bar{t} + c\bar{c}$ production via initial-state gluon radiation where both c-quarks form a jet each; (b) $t\bar{t}+c\bar{c}$ production via initial-state gluon radiation where the two c-quarks are in the same jet; (c) $t\bar{t} + 1c$ production where the c-quark originates from the initial state. Figure taken from Ref. [3].

The ATLAS Collaboration [1] at the Large Hadron Collider (LHC) [2] at CERN has now performed their first measurement of the inclusive cross-section for $t\bar{t}$ production in association with charm quarks [3] to improve the understanding of these final states. This measurement uses the full Run 2 proton–proton ATLAS dataset at $\sqrt{s} = 13$ TeV taken between 2015 and 2018, corresponding to an integrated luminosity of $140\,\text{fb}^{-1}$. The measurement selects $t\bar{t}$ events with one or two electrons or muons and at least one additional jet in the final state. A custom flavour-tagging algorithm is employed to simultaneously identify b-jets and c-jets. For the first time, this measurement then uses regions sensitive to the presence of c-quarks to extract separate fiducial cross-sections for $t\bar{t}+{\geq}2c$ and $t\bar{t}+1c$ production. The former includes events with at least two c-jets, targeting primarily the $t\bar{t} + c\bar{c}$ process illustrated in Figure 1a, while the latter comprises events with one additional c-jet, originating from either the $t\bar{t} + 1C$ scenario (Figure 1b), from $t\bar{t}+1c$ production (Figure 1c), or from $t\bar{t}+c\bar{c}$ where one of the c-jets falls outside the acceptance phase space of the detector. The CMS Collaboration had previously performed the first measurement of $t\bar{t} + c\bar{c}$ cross-sections and found them consistent with SM $t\bar{t}$ simulation predictions at NLO accuracy in QCD [4].

This document and the associated talk summarise the methodology and results of the new ATLAS measurement [3]. Detailed references to all methods, calculations, software tools, etc. are skipped here for brevity but can be found in the original publication.

2   Simulation of signal and background

One key component in scrutinising $t\bar{t}$ production in association with heavy-flavour jets is the modelling of these processes through Monte Carlo (MC) simulations. Recent ATLAS measurements of $t\bar{t}+b\bar{b}$ and $t\bar{t}H(H \rightarrow b\bar{b})$ production [5,6] have found the modelling of $t\bar{t}+{\geq}1c$ to be a limiting factor in the precision of the measurements. In the absence of dedicated $t\bar{t}+c\bar{c}$ computations, inclusive $t\bar{t}$+jets simulations at NLO accuracy interfaced to a parton shower (PS) (NLO+PS) provide the best available model for $t\bar{t}+{\geq}1c$. Here, the gluon splittings into $c\bar{c}$ pairs are modelled through the PS algorithm, not in the matrix-element calculation. The $t\bar{t} + {\geq}1b$ process, on the other hand, can be modelled through $t\bar{t} + b\bar{b}$ four-flavor scheme (4FS) matrix elements, where the b-quark is treated as a massive particle and the production of the $b\bar{b}$ pair is described directly in the calculation of the hard interaction.

In the ATLAS measurement, $t\bar{t}$+jets simulations in the five-flavor scheme (5FS), generated with POWHEG BOX 2 [7–10] and the NNPDF3.0NLO 5FS parton distribution function (PDF) set [11], are used to model the $t\bar{t} + {\geq}1c$ and $t\bar{t} +$ light processes, where the latter refers to $t\bar{t}$

events with additional jets not originating from b-quarks or c-quarks. $t\bar{t} + b\bar{b}$ 4FS simulations using POWHEG BOX RES [12], OPENLOOPS [13–15] and the NNPDF3.0NLO 4FS PDF set [11] are employed for the $t\bar{t} + {\geq}1b$ process. Both are interfaced to PYTHIA 8 [16] for the simulation of the PS and hadronisation using the ATLAS A14 set of tuned parameters [17] and the NNPDF2.3LO PDF set [18]. Reweightings and alternative setups (POWHEG BOX 2+HERWIG 7 [19–21], varied NLO matching) are employed to assess modelling uncertainties. For comparison only, an additional set of $t\bar{t}$+jets events is generated using MADGRAPH5_AMC@NLO [22] interfaced to HERWIG 7.

All simulated events are processed through the full ATLAS detector simulation and reconstruction chain, and are further split into four categories: $t\bar{t} + {\geq}2c$, $t\bar{t} + 1c$, $t\bar{t} + {\geq}1b$, and $t\bar{t} + \text{light}$. The categorisation is based on the presence of additional b-jets and c-jets clustered at the stable particle level before detector simulation. Those events with one or more b-jets are assigned to the $t\bar{t} + {\geq}1b$ category, those with one c-jet to the $t\bar{t} + 1c$ category, and those with at least two c-jets to the $t\bar{t} + {\geq}2c$ category. The latter two classes of events must not contain any additional b-jets. All remaining events are assigned to the $t\bar{t} + \text{light}$ category. The categorisation is used to combine the $t\bar{t}$+jets and $t\bar{t} + b\bar{b}$ simulations by removing all $t\bar{t} + {\geq}1b$ events from the 5FS $t\bar{t}$+jets sample and all but the $t\bar{t} + {\geq}1b$ events from the 4FS $t\bar{t} + b\bar{b}$ sample, avoiding double-counting.

Various processes are considered as background to the measurement, including single-top-quark production, associated production of single top quarks or top-quark pairs with a vector boson, and W+jets, Z+jets and diboson production. Dedicated MC simulations are used to model these processes. Background contributions from events with fake-lepton signatures, either originating from non-prompt leptons or from hadronic activity misidentified as leptons, are estimated using data-driven methods in the single-lepton channel and from MC simulations in the dilepton channel. Uncertainties in the $t\bar{t} + {\geq}1b$, $t\bar{t} + \text{light}$ and single-top-quark background processes are assessed through dedicated modelling uncertainties, while normalisation uncertainties are assigned to the minor backgrounds.

3 Analysis strategy

A dedicated flavour-tagging algorithm, the b/c-tagger, was developed for this analysis to simultaneously identify b-jets and c-jets with high efficiency. While the standard ATLAS DL1r b-tagging algorithm [23] provides high efficiency for b-jets, the lack of c-tagging working points mandates this custom approach. The output scores provided by DL1r were re-optimised into a two-dimensional discriminant defining five b/c-tagger bins: two each for c-jet (c@11%, c@22%) and b-jet (b@60%, b@70%) tagging, and one for untagged jets, as illustrated in Figure 2. When using the looser tagging bins, the tighter bins are implicitly included, e.g., the c@22% bin includes all jets that pass c@11%.

The analysis uses a suite of single-electron and single-muon triggers to select events with one or two electrons (e) or muons (μ), with one required to satisfy the trigger conditions. The single-lepton channel selects one lepton and ≥ 5 jets, while the dilepton channel selects two leptons and ≥ 3 jets, with at least three jets passing b@70% or c@22% criteria in either channel. Same-flavour lepton pairs are vetoed for low invariant masses and around the Z boson mass. The single-lepton channel is split into events with five jets and events with six or more jets (5-jet-exclusive and 6-jet-inclusive). The dilepton channel is split into 3-jet-exclusive and 4-jet-inclusive events.

The five b/c-tagger bins define 19 orthogonal control regions (CRs) and signal regions (SRs) across the single-lepton and dilepton channels, providing sensitivity to $t\bar{t} + {\geq}2c$, $t\bar{t} + 1c$, and controlling background contributions. The twelve CRs are designed to either exclude
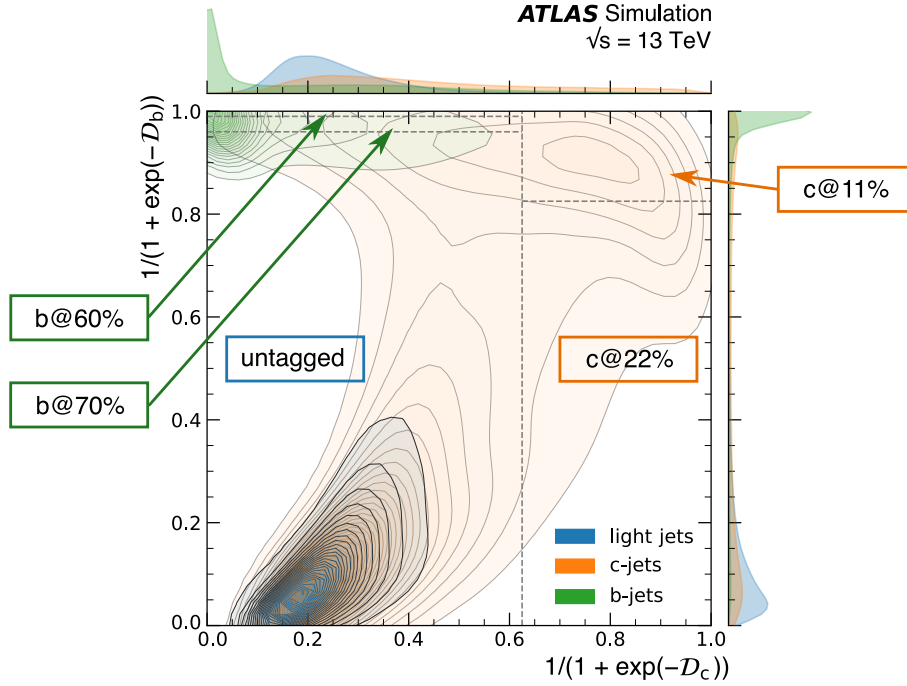
Figure 2: Distribution of light, c-, and b-jets for the two-dimensional b/c-tagger in simulated $t\bar{t}$ events. Dashed lines correspond to the edges of the five b/c-tagger bins. For visualisation purposes a standard logistic function is applied to both axes of the discriminant. Figure taken from Ref. [3].

c-tagged jets or select events with exactly one c-tagged jet, isolating $t\bar{t}+\geq 1b$, $t\bar{t}+$ light, and other background contributions. The seven SRs require at least two (one) jet tagged with c@22% in the single-lepton (dilepton) channel and have different predicted purities of $t\bar{t}+\geq 2c$ and $t\bar{t}+1c$ events. These regions are then used to extract the $t\bar{t}+\geq 2c$ and $t\bar{t}+1c$ signal strengths – defined as the ratios of observed yields to the SM expectations – from the data using a profile likelihood fit. Additionally, normalisation factors for the $t\bar{t}+\geq 1b$ and $t\bar{t}+$ light background contributions are included in the fit and are treated as free-floating parameters. Systematic uncertainties relating to the modelling of the signal and background processes, the dedicated flavour-tagging algorithm, and all other instrumental uncertainties are included in the fit as nuisance parameters constrained by Gaussian penalty terms. While the CRs are included as a single bin in the fit, the jet-inclusive and jet-exclusive SRs use jet-multiplicity and invariant-mass distributions of the c-tagged jets, respectively.

The measurements are performed in a fiducial phase space defined at the stable particle level that mimics the selection criteria of the analysis. Apart from the lepton requirements, the fiducial phase space requires at least one additional jet, but no explicit requirements are imposed on its flavour. The measurements are also performed in a more inclusive volume without requirements on the $t\bar{t}$ decay products and the jet multiplicity.

4   Results

Good post-fit agreement between data and simulation is observed in all CRs and SRs of the analysis. The goodness of fit was evaluated using a *saturated model* and the compatibility with data was found to be 98 %. Using the extracted $t\bar{t}+\geq 2c$ and $t\bar{t}+1c$ signal strengths, the
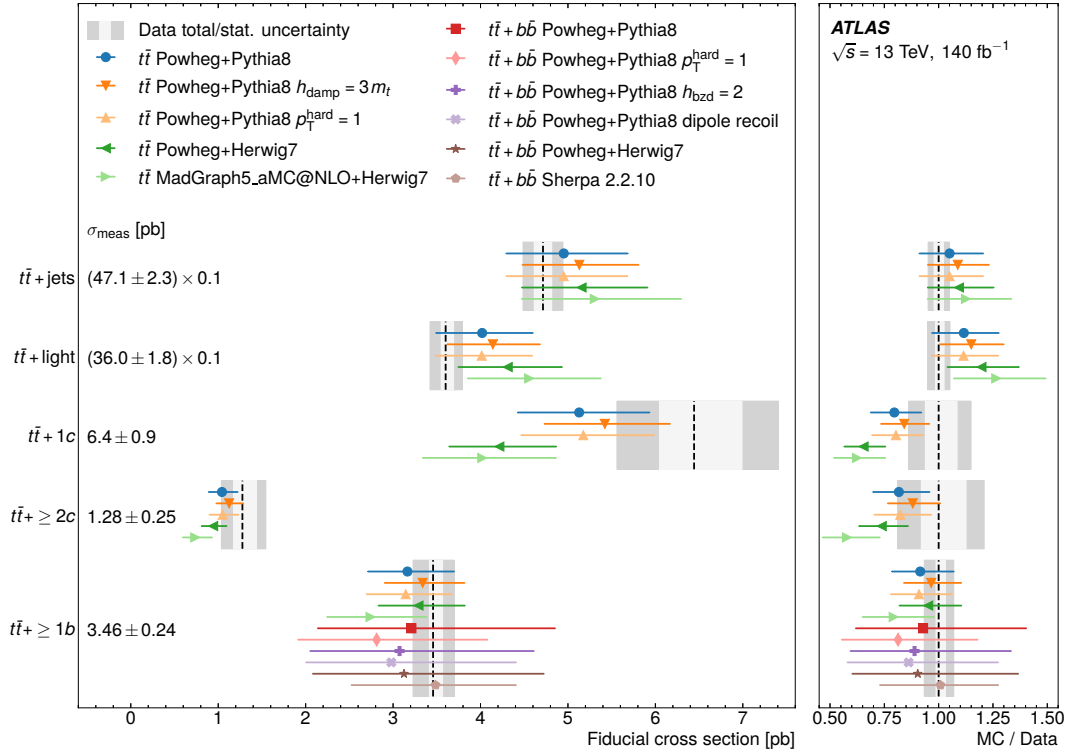
Figure 3: Measured fiducial cross-section values in comparison with various NLO+PS predictions from inclusive $t\bar{t}$+jets and $t\bar{t}+b\bar{b}$ simulations. Figure taken from Ref. [3].

fiducial cross-sections are determined to be

$$\sigma^{\text{fid}}(t\bar{t} + \geq 2c) = 1.28\,^{+0.16}_{-0.10}\,(\text{stat})\,^{+0.21}_{-0.22}\,(\text{syst})\,\text{pb} = 1.28\,^{+0.27}_{-0.24}\,\text{pb}, \tag{1}$$

$$\sigma^{\text{fid}}(t\bar{t} + 1c) = 6.4\,^{+0.5}_{-0.4}\,(\text{stat})\,\pm 0.8\,(\text{syst})\,\text{pb} = 6.4\,^{+1.0}_{-0.9}\,\text{pb}. \tag{2}$$

The largest contributing sources of uncertainties are the modelling of $t\bar{t} + \geq 1c$, $t\bar{t} + \geq 1b$, and $t\bar{t}$ + light, in particular in the NLO matching and the PS, the uncertainties in the b/c-tagger, and the data statistics. The measured $t\bar{t} + \geq 1b$ normalisation factor is compatible with those obtained in the dedicated ATLAS $t\bar{t} + b\bar{b}$ measurements [5].

The measured values are compared to various NLO+PS predictions from inclusive $t\bar{t}$+jets and $t\bar{t}+b\bar{b}$ simulations in Figure 3. The prediction for $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ are largely consistent with the measured values, though all predictions underpredict the observed values. Considering scale and PDF variations on the predictions, the $t\bar{t}$ POWHEG BOX 2+PYTHIA 8 predictions agree within 1.1 standard deviations of measurement and prediction uncertainties. Agreement between 1.2 and 2.0 standard deviations is observed for the POWHEG BOX 2+HERWIG 7 and the MADGRAPH5_AMC@NLO+HERWIG 7 predictions. Stability tests, including combined and independent $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ parameterisations, validate the robustness of the fit.

In addition to the fiducial cross-sections, the ratios of $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ to total $t\bar{t}$+jets production are extracted in the fiducial and the more inclusive volume. The values were found to be $R^{\text{inc}}_{t\bar{t}+\geq 2c} = (1.23 \pm 0.25)\%$ and $R^{\text{inc}}_{t\bar{t}+1c} = (8.8 \pm 1.3)\%$ in inclusive volume, in agreement with NLO+PS simulations at a similar level as the fiducial cross-sections.

5 Conclusion

The ATLAS Collaboration at the LHC has performed their first measurement of the inclusive cross-section for $t\bar{t}$ production in association with charm quarks [3]. The measurement uses the full Run 2 proton–proton collision data sample at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of $140\,\mathrm{fb}^{-1}$, collected between 2015 and 2018. Using single-lepton and dilepton events and a custom flavour-tagging algorithm to identify b-jets and c-jets, fiducial cross-sections for $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ production are extracted separately from the data for the first time. The measurements are found to be in agreement with various NLO+PS predictions from inclusive $t\bar{t}$+jets and $t\bar{t} + b\bar{b}$ simulations, though all predictions underpredict the observed values. Precision is limited by modelling uncertainties ($t\bar{t} + \geq 1c$, $t\bar{t} + \geq 1b$, $t\bar{t}$ +light), uncertainties in the b/c-tagger, and data statistics. Measurements of the ratios of the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ cross-sections to the inclusive $t\bar{t}$+jets cross-section are also performed and found to be in agreement with the predictions from NLO+PS simulations at a similar level as the fiducial cross-sections.

Acknowledgements

The author received support from the U.S. Department of Energy Award No. DE-SC0007881.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008), doi:10.1088/1748-0221/3/08/S08003.

[2] L. Evans and P. Bryant, *LHC Machine*, JINST **3**, S08001 (2008), doi:10.1088/1748-0221/3/08/S08001.

[3] ATLAS Collaboration, *Measurement of top-quark pair production in association with charm quarks in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **860**, 139177 (2025), doi:10.1016/j.physletb.2024.139177, 2409.11305.

[4] CMS Collaboration, *First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **820**, 136565 (2021), doi:10.1016/j.physletb.2021.136565, 2012.09225.

[5] ATLAS Collaboration, *Measurement of $t\bar{t}$ production in association with additional b-jets in the eμ final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **01**, 068 (2025), doi:10.1007/JHEP01(2025)068, 2407.13473.

[6] ATLAS Collaboration, *Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **85**(2), 210 (2025), doi:10.1140/epjc/s10052-025-13740-x, 2407.10904.

[7] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09**, 126 (2007), doi:10.1088/1126-6708/2007/09/126, 0707.3088.

[8] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11**, 040 (2004), doi:10.1088/1126-6708/2004/11/040, hep-ph/0409146.

[9] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11**, 070 (2007), doi:10.1088/1126-6708/2007/11/070, 0709.2092.

[10] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06**, 043 (2010), doi:10.1007/JHEP06(2010)043, 1002.2581.

[11] NNPDF Collaboration, R. D. Ball *et al.*, *Parton distributions for the LHC run II*, JHEP **04**, 040 (2015), doi:10.1007/JHEP04(2015)040, 1410.8849.

[12] T. Ježo, J. M. Lindert, N. Moretti and S. Pozzorini, *New NLOPS predictions for $t\bar{t} + b$-jet production at the LHC*, Eur. Phys. J. C **78**(6), 502 (2018), doi:10.1140/epjc/s10052-018-5956-0, 1802.00426.

[13] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108**, 111601 (2012), doi:10.1103/PhysRevLett.108.111601, 1111.5206.

[14] A. Denner, S. Dittmaier and L. Hofer, COLLIER: *A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212**, 220 (2017), doi:10.1016/j.cpc.2016.10.013, 1604.06792.

[15] F. Buccioni, J.-N. Lang, J. M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang and M. F. Zoller, *OpenLoops 2*, Eur. Phys. J. C **79**(10), 866 (2019), doi:10.1140/epjc/s10052-019-7306-2, 1907.13071.

[16] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015), doi:10.1016/j.cpc.2015.01.024, 1410.3012.

[17] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014).

[18] NNPDF Collaboration, R. D. Ball *et al.*, *Parton distributions with LHC data*, Nucl. Phys. B **867**, 244 (2013), doi:10.1016/j.nuclphysb.2012.10.003, 1207.1303.

[19] M. Bähr *et al.*, *Herwig++ physics and manual*, Eur. Phys. J. C **58**, 639 (2008), doi:10.1140/epjc/s10052-008-0798-9, 0803.0883.

[20] J. Bellm *et al.*, *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76**(4), 196 (2016), doi:10.1140/epjc/s10052-016-4018-8, 1512.01178.

[21] J. Bellm *et al.*, *Herwig 7.1 Release Note* (2017), 1705.06919.

[22] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07**, 079 (2014), doi:10.1007/JHEP07(2014)079, 1405.0301.

[23] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83**, 681 (2023), doi:10.1140/epjc/s10052-023-11699-1, 2211.16345.