**THE EUROPEAN
PHYSICAL JOURNAL C**

Searches for new physics with boosted top quarks in the MadAnalysis 5 and Rivet frameworks

Jack Y. Araz[1,a] ⓘ, **Andy Buckley**[2,b] ⓘ, **Benjamin Fuks**[3,c] ⓘ

[1] Institute for Particle Physics Phenomenology, Durham University, Durham, UK

[2] School of Physics and Astronomy, University of Glasgow, Glasgow, UK

[3] Laboratoire de Physique Théorique et Hautes Énergies (LPTHE), UMR 7589, Sorbonne Université et CNRS, 4 place Jussieu, 75252 Paris Cedex 05, France

Abstract High-momentum top quarks are a natural physical system in collider experiments for testing models of new physics, and jet substructure methods are key both to exploiting their largest decay mode and to assuaging resolution difficulties as the boosted system becomes increasingly collimated in the detector. To be used in new-physics interpretation studies, it is crucial that related methods get implemented in analysis frameworks allowing for the reinterpretation of the results of the LHC such as MADANALYSIS 5 and RIVET. We describe the implementation of the HEPTOPTAGGER algorithm in these two frameworks, and we exemplify the usage of the resulting functionalities to explore the sensitivity of boosted top reconstruction performance to new physics contributions from the Standard Model Effective Field Theory. The results of this study lead to important conclusions about the implicit assumption of Standard-Model-like top quark decays in associated collider analyses, and for the prospects to constrain the Standard Model Effective Field Theory via kinematic observables built from boosted semi-leptonic $t\bar{t}$ events selected using HEPTOPTAGGER.

1 Introduction

Since the resurrection of jet-substructure methods as probes for new particles at the LHC [1,2], boosted topologies in which multiple decay products from heavy intermediate states fall into a single large-radius (large-R) jet have seen wide application in searches for new physics [3–8]. While not initially considered in the early days of the LHC, these jet substructure techniques are now indeed largely used to extend the sensitivity of searches for new physics. This is particularly the case as the currently null results of those searches indicate that any relevant physics beyond the Standard Model (BSM) is most-likely located at a large mass scale, featuring heavy particles whose production and decay would naturally yield highly-boosted lighter Standard Model (SM) objects.

Many collider signatures can benefit from the usage of jet substructure methods, as they can be generally applied to tag many SM and BSM particles when they are produced with a high Lorentz boost. Among these, the top quark is an important target, for two reasons. First, the top quark is the highest-mass fermion in the SM, featuring a Yukawa coupling value close to 1. This makes it a natural candidate to provide an explanation for the hierarchy problem, and to play the role of a mediator that couples to new-physics sectors (e.g. through the Higgs field). Second, boosted methods can provide better background-rejection power than a classic 'resolved' reconstruction of the top quark kinematics. As a high-mass, colour-charged and non-hadronising particle, the top quark is the most complex SM resonance to reconstruct from fully resolved decay components. This not only requires highly performant b-tagging, but also suffers from either a complicated lepton and missing-momentum reconstruction or the resolution difficulties inherent to reconstructing a fully hadronic $b\bar{q}q'$ final state.

Jet-substructure methods offer a way to bypass many of the difficulties related to the reconstruction and identification of hadronically-decaying top quarks by relying on one single large-radius jet in place of three small-radius ones. In addition, such an option generally exploits the presence of *two* heavy SM particles' decay hierarchies within the large-R jet (the top quark itself and the W boson originating from its

[a] e-mail: jack.araz@durham.ac.uk

[b] e-mail: andy.buckley@glasgow.ac.uk

[c] e-mail: fuks@lpthe.jussieu.fr (corresponding author)

decay), together with information on the internal momentum and angular structure of all jet constituents (with or without b-tagging requirements) to disambiguate boosted top quarks from jets originating from pure QCD background processes. A prominent tool in such studies is the HEPTOPTAGGER method [9,10], which pioneered this approach and has since gone through several rounds of enhancement such as use of variable-radius jet clustering. In the meantime more sophisticated and efficient top tagging methods have been developed. Typical examples are based on a classification of jets making use of the radiation pattern within a jet (also known as shower deconstruction) [11], on advanced machine learning techniques (we refer to Ref. [12] for an overview) relying on observables like the jet transverse momentum and mass, the dispersion of its constituents estimated through N-subjettiness variables [13,14], splitting scales [15], energy correlation functions [16,17], as well as on jet image analysis by means of neutral networks [18–20] and image or language recognition techniques [21–24]. More recently, a series of machine-learning methods embedding Lorentz invariance [25,26] have additionally been proposed and explored. The HEPTOPTAGGER method, however, still plays the role of being an important benchmark in the top-tagging landscape, especially in the context of use by the LHC experiments [27,28]. On the other hand, the related code has historically been unavailable for use in analysis prototyping and preservation within the two public analysis frameworks MADANALYSIS 5 [29–31] and RIVET [32,33], that are widely used across the high-energy physics community. The goal of the present work is to fill this gap, and to document, through a few examples, its addition to both frameworks. It also therefore serves as a prototype interface for integration of C++ versions of machine-learning taggers into these public analysis toolkits.

While most applications of boosted top quark reconstruction have been aimed at direct searches for new physics, the lack of tangible evidence for new high-mass resonances urges complementary studies of indirect routes through which BSM physics can manifest. A leading approach in this is that of the Standard Model Effective Field Theory (SMEFT), in which the explicit microscopic physics of a particular BSM model is replaced by an infinite set of higher-dimensional operators involving the SM fields and compatible with the SM symmetries [34–36]. The SMEFT is then an expansion in an energy scale Λ above which the effective theory breaks down and real new physics resides, so that new fields with masses comparable with Λ must be explicitly added to the model's Lagrangian. Details about the UV theory are encoded in the Wilson coefficients multiplying each operator, and the relevance of a specific new interaction is dictated (a priori) by the dimension of the corresponding operator (that is thus suppressed by some power of the effective scale). At dimension six, 84 (3045) parameters encode the leading BSM effects, assuming a flavour-blind (flavour-general)

setup [37,38]. Constraints are typically made primarily in the model-independent space of the corresponding Wilson coefficients by investigating the possibility of small and often subtle deviations from the SM expectation. Among all operators, about twenty of them impact top physics under the simplifying assumption that new physics couples dominantly to bosons and to the left-handed doublet and right-handed up-type singlet of third generation quarks [39].

Global SMEFT interpretations of measurements at the LHC in the top sector have recently been achieved by several groups [40–45]. These studies demonstrated in particular that dozens of SMEFT operators could be constrained (and therefore determined) simultaneously, correlating sometimes information originating from different sectors. It is nevertheless well known that signatures of processes involving boosted top quarks could be crucially relevant [46]. These indeed involve large momentum transfers, so that they are expected to exhibit the largest sensitivity to new physics effects in the SMEFT, and subsequently show the most sensitivity to BSM phenomena. It is therefore natural to focus on high-momentum collider event categories involving the production of boosted top quarks, and to consider them as a promising avenue to statistically constrain the viable space of Wilson coefficients associated with top quark operators.

In the present study, we make use of the HEPTOPTAGGER functionalities that we implemented in the RIVET and MADANALYSIS 5 frameworks (together the possibility of computing emulated reconstruction-level observables) to study the sensitivity of the LHC to top-related SMEFT operators, focusing on the production of a pair of boosted top quarks. However, the HEPTOPTAGGER algorithm is designed to exploit as much as possible the kinematics of the SM decay of a boosted top quark. This leads to the open question about how new physics effects arising from the introduction of nonzero top quark SMEFT operators could modify these kinematics, and hence impact the performance of HEPTOPTAGGER and, by inference, of any similar reconstruction method based on the topology arising from an SM top quark decay. As a straightforward application and keeping this in mind, we highlight important resulting issues for BSM interpretations.

In Sect. 2, we detail our technical developments in RIVET (Sect. 2.2) and MADANALYSIS 5 (Sect. 2.3), and briefly explain how to use the codes for physics studies. In Sect. 3, we exemplify the usage of these developments to estimate the impact of new physics via effective SMEFT operators on HEPTOPTAGGER performance, and how this affects the sensitivity of the present and future runs of the LHC (assuming an integrated luminosity of 300 fb^{-1} and 3000 fb^{-1} and varied levels of systematic errors) to these operators. We summarise our work in Sect. 4.

2 HEPTopTagger implementation in the Rivet and MadAnalysis 5 frameworks

2.1 Generalities

In its initial proposal [9], the HEPTopTagger algorithm is a purely deterministic top-tagging method in which boosted top reconstruction is solely achieved from the geometrical structure and properties of the constituents of a fat jet. It first defines a fat jet collection from an event final state by using the Cambridge-Aachen jet algorithm [47–49]. A procedure is next applied to all jets included in this collection, in order to decide whether they should be top-tagged.

In practice, each reconstructed fat jet is decomposed into several subjets by applying a mass drop criterion [50]. More precisely, jet clustering is iteratively undone so that each fat jet is split in two subjets, and the subjet with the smallest invariant mass is kept only if its invariant mass is large enough. Each resulting subjet is further decomposed in the same manner provided that its invariant mass is larger than some threshold. All possible triplets of jets belonging to the subjet collection obtained in this way are then filtered, and the five hardest filtered subjets are selected for boosted top quark reconstruction.

These five subjets are reclustered into three subjets, that are thus assumed to originate from a top quark decay. Events are at this stage rejected if they do not include any resulting triplet with an invariant mass that is compatible with the top mass. Top tagging stems from several requirements that are imposed on the invariant masses of the different dijet pairs that could be formed from the three subjets of any boosted top quark candidate, in particular in order to ensure the compatibility with the presence of an intermediate W boson. Moreover the transverse momentum of the top candidate is required to be at least 200 GeV.

We refer to the original documentation [9,10] for a more comprehensive and quantitative presentation of the HEPTopTagger algorithm.

The performance of any top quark tagger can be improved by using an increased set of input variables (as in most multi-variate methods), for which the explicit choices are made through a tuning process relative to a given reference. To this end, the HEPTopTagger method has been updated and now includes a variety of features that enhance the tagging efficiency and reduce the associated mistagging rates: it uses substructure mass-drop conditions [50], jet trimming [51] and pruning [4,52] algorithms, and filtering steps [2], in addition to the core requirement that the large-radius jet demonstrates the three-pronged structure characteristic of a boosted top-quark's hadronic decay. In the current version of the HEPTopTagger package, all these methods are used together in a multi-variate classification [53,54] which maximises the expected tagging performance.

Access to this tool and all its embedded features within public frameworks like MADANALYSIS 5 or RIVET is thus crucial for prototyping and reproducing collider-event data analyses, a key activity in collider phenomenology. In the rest of this section, we discuss technical details about the embedding of HEPTOPTAGGER in these two software tools, and describe how they could practically be used. In practice, we rely on the latest public version of HEPTOPTAGGER (i.e. its version 2 available from the webpage https://www.thphys.uni-heidelberg.de/~plehn/index.php?show=heptoptagger). Moreover, we have validated our implementations by confronting the results of a few test calculations obtained by using the two interfaced versions of HEPTOPTAGGER to those returned by HEPTOPTAGGER when used in a standalone mode.

2.2 Jet substructure tools in RIVET

The implementation of HEPTOPTAGGER within RIVET has been designed on top of its existing jet-analysis toolkit, using the 'smearing projection' machinery that simulates kinematic and particle-identification misreconstruction through transfer functions, while preserving links between particle-level and reconstruction-level physics objects. When jet substructure methods are involved, dedicated smearing methods are required, as many observables (e.g. N-subjettinesses) are sensitive to angular correlations between the jet constituents. It is therefore necessary to model the detector's finite angular resolution to get a realistic detector response, including the inefficiencies related to the hadronic calorimeter. This is achieved, as detailed in Ref. [55], through the directional smearing of the pseudo-rapidity η and azimuthal angle φ variables defining the direction of every jet constituent. As angular deflections are more significant for constituents with a low transverse momentum p_T, this smearing is made p_T-dependent with greater angular stability at higher momentum. The specific form used, known to describe jet-substructure effects well on the public data, is angular smearing by a Gaussian with a mean of zero and a standard deviation given by

$$\sigma_{\text{ang}} = \frac{\alpha}{1 + e^{\beta(p_T^i - \gamma)}} \,. \tag{1}$$

Here α, β and γ are free parameters, set to 0.045, 0.013 and 31.15, respectively, from the fit detailed in Ref. [55]. Additionally, energy-resolution smearing was performed, using relative scaling by a Gaussian with mean of 1, and width $\sigma_E \sim 10\%$.

Our implementation of the HEPTOPTAGGER method in RIVET relies on an object of the `HTT` class, normally to be declared as a member variable of an analysis (or projection) class,

```
HTT _tagger;
```

Table 1 Accessors equipping the HEPTOPTAGGER wrapper implemented in RIVET

Accessor	Functionality
`const Jet topJet()`	Top quark candidate (returned as a `Jet` object)
`const Jet bJet()`	b-jet candidate (returned as a `Jet` object)
`const Jet wJet()`	Combined subjets compatible with a W-boson candidate (returned as a `Jet` object)
`const Jet w1Jet()`	Leading subjet constituting a W-boson candidate (returned as a `Jet` object)
`const Jet w2Jet()`	Sub-leading subjet constituting a W-boson candidate (returned as a `Jet` object)
`const PseudoJet topJet()`	Top quark candidate (returned as a `PseudoJet` object)
`const PseudoJet bJet()`	b-jet candidate (returned as a `PseudoJet` object)
`const PseudoJet wJet()`	Combined subjets compatible with a W-boson candidate (returned as a `PseudoJet` object)
`const PseudoJet w1Jet()`	Leading subjet constituting a W-boson candidate (returned as a `PseudoJet` object)
`const PseudoJet w2Jet()`	Sub-leading subjet constituting a W-boson candidate (returned as a `PseudoJet` object)
`Jets subjets()`	p_{T}-ordered subjets
`double prunedMass()`	Pruned top quark mass
`double unfilteredMass()`	Mass of the triplet of subjets after unclustering, and before filtering
`double deltaTopMass()`	Difference between the reconstructed top mass and the true top mass $|m_{\mathrm{rec}} - m_t|$
`bool isTopTagged()`	Boolean indicating if the top jet has a mass compatible with the top mass, satisfies two-dimensional mass plane requirements, and has a p_{T} above some threshold
`bool passedMassCutTop()`	Boolean indicating if the top jet has a mass compatible with the top mass
`bool passedMassCut2D()`	Boolean indicating if the top jet satisfies two-dimensional mass plane requirements

The `HTT` class is defined in the header file ``Rivet/Tools /RivetHTT.hh''. All available parameters for this wrapper are initialised through an `HTT::InputParameters` object, that can be used for any further modification relevant for the needs of the user. A simple example is

```
HTT::InputParameters parameters;
parameters.mass_drop = 0.8; // mass
    drop rate
parameters.filt_N = 5; // nr of
    filtered subjets
_tagger.setParams(parameters);
```

The list of available parameters can be found in the definition of the C++ structure `HTT::InputParameters` in the file ``Rivet/Tools/RivetHTT.hh''. All parameters that are not explicitly initialised by the user keep their default values which have been chosen according to Refs. [9,10]. During the execution of a RIVET analysis, a reclustered jet, instantiated as a `Jet` object, can be processed by the methods of the `_tagger` object, for example in

```
_tagger.calc(fatjets[0]);
```

Here, `fatjets[0]` refers to the leading (i.e. highest-p_{T}) jet included in a vector of clustered jets called `fatjets` (the object `fatjets` being thus of type `Jets`). The computation yields the creation of various accessors that returns a variety of information into native RIVET objects. This list of accessors is shown in Table 1.

For a practical example, we refer to the illustrative analysis that can be found in RIVET's ``analyses/examples/ EXAMPLE_HTT.cc'' file.

2.3 Jet substructure tools in MADANALYSIS 5

Since 2021, MADANALYSIS 5 and its SFS framework for fast simulation of detector effects [56] have been equipped with jet substructure tools and methods.[1] In particular, the smearing functionality implemented in the SFS framework allows for modifications of the properties of the jets' constituents, so that the SFS is suitable for the embedding of HEPTOP-TAGGER in a way similar to what was achieved for RIVET in Sect. 2.2. As the substructure branch is so far largely undocumented, we take benefit from the present work to provide some details on its functioning and how to make use of the code to embed top tagging in a generic analysis.

When a jet reconstruction algorithm is turned on in MAD-ANALYSIS 5, a so-called '*primary*' jet collection is built from a hadronised event. This primary jet collection is equivalent to the sole jet collection that used to be built in versions 1.X.Y of the code, which was documented in [31,56]. In practice, the code makes use of its interface with FASTJET [57], that can be turned on from the MADANALYSIS 5 command line interface by typing

```
set main.fastsim.package = fastjet
```

A specific jet algorithm is then activated through the commands

---

[1] See the GITHUB branch https://github.com/MadAnalysis/ madanalysis5/tree/substructure, as well as the beta versions of MAD-ANALYSIS 5 v2.0.Z (available at https://github.com/MadAnalysis/ madanalysis5/releases).

```
set main.fastsim.algorithm = \
<algorithm>
set main.fastsim.<property> = <value>
```

The list of supported algorithms, together with the available properties, is provided in [56]. By default, the anti-k_T jet algorithm [58] is considered, with a radius parameter $R = 0.4$ (`radius`) and a minimum p_T value of 5 GeV (`ptmin`). The primary jet collection is identified by its jet identifier (or `JetID`), that is fixed to `Ma5Jet` by default. This identifier can be further modified through the command

```
set main.fastsim.JetID = <new JetID>
```

Additional jet collections can be instantiated through

```
define jet_algorithm <JetID> <algo>\
    [<parameters>]
```

where `<JetID>` refers to the identifier of the collection, `<algo>` to the associated clustering algorithm, and where any algorithm-specific parameter can be optionally fixed through comma-separated or space-separated equalities (otherwise default values are used). For instance, typing

```
define jet_algorithm CA08 cambridge \
    radius=0.8  tmin=200
```

defines a jet collection coined `CA08`, in which jets are reconstructed by means of the Cambridge-Aachen jet algorithm [47–49], with a radius parameter set to 0.8 and a minimum p_T value of 200 GeV. Parameters can also be altered through specific commands, like for instance in

```
set CA08.radius = 0.8
set CA08.ptmin = 200
```

Once multiple jet collections are defined, constituent-based smearing is always applied to the properties of all final-state hadrons before the different reconstructions are performed. This contrasts with the setup in which a single collection is defined, as here users can decide to smear reconstructed objects instead of their constituents. Reconstruction efficiencies can also be provided from the command line interface (see [56]), but they will only be applied to the primary jet collection. This limiting behaviour can however be bypassed by employing the expert mode of the code, in which users implement their analysis directly in C++ (and are thus free to do whatever they want). We therefore focus only on this expert mode from now on.[2]

At the level of the C++ code generated by MADANALYSIS 5 (or implemented from scratch by expert users), the primary jet collection can be accessed through the standard accessor `event.rec()->jets()` (as described in [30,31]), and all jet collections (including the primary one)

can be accessed through the accessor `event.rec()->jets(<JetID>)` (with `<JetID>` being the identifier referring to the collection). These accessors return a vector of pointers to constant `RecJetFormat` objects (or `RecJet` objects for short), the entire vector being also of the shorthand type `RecJets`.

In the version 2.0.X of MADANALYSIS 5, a `Substructure` namespace has been implemented and includes wrappers to a large set of FASTJET and FASTJET CONTRIB functionalities. This substructure module allows for three standard infrared and collinear safe jet-clustering algorithms, that can be initialised as for instance through

```
Substructure::Cluster cluster;
cluster.Initialize(Substructure::
    antikt, 0.4, 20., isExclusive =
    false);
```

This initialises a `Cluster` object named `cluster` in which jet reconstruction relies on the anti-k_T algorithm with parameter $R = 0.4$, and that selects reconstructed jets featuring $p_T > 20$ GeV. In order to make use of the Cambridge-Aachen or the generalised k_T [57] algorithm, the first argument of the `Initialize` method needs to be set to `Substructure::cambridge` and `Substructure::kt` respectively. The next arguments are related to the two options available for the three supported algorithms (namely the radius parameter R and the minimum p_T requirement applied on the reconstructed jets), and the last optional argument (`isExclusive`) indicates whether leptons and photons originating from hadron decays have to be included in their respective collections in addition to be considered as jet constituents (`isExclusive = false`), or not (`isExclusive = true`). Next, clustering is executed through the command

```
cluster.Execute(<event>, <JetID>);
```

where `<JetID>` is the identifier of the jet collection to use to store the output of the clustering, and `<event>` is an `EventFormat` object pointing to the whole event. Smearing and reconstruction efficiencies are automatically included, if provided by the user (see Ref. [56]).

Clustered jets can be further manipulated, either one by one or all together. For instance, the first of the following commands defines a new collection `FilteredJets` as a subselection of all reconstructed (primary) jets satisfying $p_T > 20$ GeV and $|\eta| < 2.5$. The next two lines are dedicated to the initialisation of a new clustering method (the Cambridge-Aachen algorithm with a radius parameter $R = 0.5$, that is the sole parameter that can be specified here), with which those jets will be reclustered,

```
RecJets FilteredJets = filter(event.
    rec()->jets(), 20.0, 2.5);
Substructure::Recluster recluster;
recluster.Initialize(Substructure::
    cambridge, 0.5);
```

---

[2] The command line interface of MADANALYSIS 5 can be used for the generation of a skeleton C++ code, that can then be modified in a second step according to the wishes of the user.

Table 2 Accessors equipping the HEPTopTagger wrapper implemented in MADANALYSIS 5

Accessor	Functionality
`const RecJet top()`	Top quark candidate (returned as a `RecJet` object)
`const RecJet b()`	b-jet candidate (returned as a `RecJet` object)
`const RecJet W()`	Combined subjets compatible with a W-boson candidate (returned as a `RecJet` object)
`const RecJet W1()`	Leading subjet constituting a W-boson candidate (returned as a `RecJet` object)
`const RecJet W2()`	Sub-leading subjet constituting a W-boson candidate (returned as a `RecJet`object)
`RecJets subjets()`	p_T-ordered subjets (returned as a vector of `RecJet` objects, also known as a `RecJets` object)
`MAfloat32 pruned_mass()`	Pruned top quark mass
`MAfloat32 unfiltered_mass()`	Mass of the triplet of subjets after unclustering, and before filtering
`MAfloat32 delta_top()`	Difference between the reconstructed top mass and the true top mass $|m_{\rm rec} - m_t|$
`MAbool is_tagged()`	Boolean indicating if the top jet has a mass compatible with the top mass, satisfies two-dimensional mass plane requirements, and has a p_T above some threshold
`MAbool is_maybe_top()`	Boolean indicating if the top jet has a mass compatible with the top mass
`MAbool is_masscut_passed()`	Boolean indicating if the top jet satisfies two-dimensional mass plane requirements

Here, we assume that the primary jets have been clustered through some (unspecified) algorithm. Next, we make use of the `Recluster` object, a first time on the whole jet collection, and a second time specifically on the leading jet,

```
ecJets ReclusteredJets = recluster.
    Execute(FilteredJets);
onst RecJet ReclusteredLeadingJet=
    recluster.Execute(FilteredJets[0]);
```

As another example, we now discuss jet reconstruction in which the radius parameter R is variable [59].[3] Such a method can be used from the `Substructure` wrapper as follows,

```
Substructure::VariableR variableR;

MAfloat32 rho=2000., minR=0., maxR
    =2., ptmin=0.0;
Substructure::VariableR::ClusterType
    clusterType =
Substructure::VariableR::AKTLIKE;
Substructure::VariableR::Strategy
    strategy =
Substructure::VariableR::Best;
MAbool isExclusive = false;

variableR.Initialize(rho, minR, maxR,
    clusterType, strategy, ptmin,
    isExclusive);
```

The clustering type must be CALIKE (Cambridge-Aachen), KTLIKE (k_T algorithm) or AKTLIKE (anti-k_T algorithm), the parameters `minR` and `maxR` stand for the minimum and maximum radius values allowed, and the internal clustering strategy to be used by FASTJET has to be among `Best`, `N2Tiled`, `N2Plain`, `NNH` or `Native`. We refer to Ref. [59] for more information. Reclustering is then proceeded as above,

```
RecJets variableRJets = variableR.Execute
    (FilteredJets);
RecJets variableRLeadingJet = variableR.
    Execute(FilteredJets[0]);
```

In order to enable the usage of HEPTOPTAGGER within MADANALYSIS 5, the package must first be downloaded and linked to the code. This is achieved by typing in the MAD-ANALYSIS 5 command line interface

```
install HEPTopTagger
```

once FASTJET and FASTJET CONTRIB are installed and available (which is achieved by typing in the interpreter the command `install fastjet`). When implementing an analysis in C++, the execution of HEPTOPTAGGER is controlled from a dedicated structure called `Substructure::HTT::InputParameters`. The latter is defined in the file "`tools/SampleAnalyzer/Interfaces/HEPTopTagger/HTT.h`", together with all associated parameters and methods, and it is documented in the file "`tools/SampleAnalyzer/Interfaces/HEPTopTagger/README.md`". Taking the example introduced in Sect. 2.2, a simple example of initialisation would read

```
Substructure::HTT::InputParameters
    parameters;
parameters.mass_drop = 0.8;
parameters.filt_N = 5;

Substructure::HTT tagger;
tagger.Initialize(parameters);
```

HEPTOPTAGGER is then executed as

```
tagger.Execute(filteredJets[0]);
```

As for the embedding into RIVET, this method leads to the generation of a variety of accessors that allows for the exploration of the properties of the would be top-jet. Their list is given in Table 2. For more detailed practical examples on the

---

[3] See the webpage https://phab.hepforge.org/source/fastjetsvn/browse/contrib/contribs/VariableR/tags/1.2.1/ for more information.

usage of jet substructure techniques and HEPTOPTAGGER within MADANALYSIS 5, we refer to the tutorial available from https://github.com/MadAnalysis/tutorial_osu.

3 Exploring new physics effects with boosted top quarks in the SMEFT

In this section, we demonstrate the use of HEPTOPTAGGER (version 2) within the RIVET and MADANALYSIS 5 frameworks, and we study the potential impact of SMEFT operators on boosted top quark decays. The set of relevant operators that we consider is introduced in Sect. 3.1. In Sect. 3.2, we focus on the production of a semi-leptonically decaying $t\bar{t}$ pair to investigate how SMEFT deviations in the properties of boosted top quarks affect the performance of top taggers (through deviations from the taggers' expectations of SM-like top quark decay properties). Next, we make use of our findings to derive in Sect. 3.3 the sensitivity of a typical analysis of boosted top-pair production and decay to various SMEFT operators poorly constrained by other means.

3.1 Theoretical framework

In the absence of any explicit evidence for new fields and interactions beyond the SM, effective field theories provide a natural path to scrutinising the impact of hypothetical BSM physics at the electroweak scale $\Lambda_{\rm EW}$. In this context, the SMEFT paradigm offers a very promising framework allowing for the exploration of heavy new physics. The SMEFT is an effective field theory expansion in an energy scale Λ that is assumed to satisfy $\Lambda \gg \Lambda_{\rm EW}$. The model Lagrangian is defined via a set $\{\mathcal{O}_1, \mathcal{O}_2, ...\}$ of higher-dimensional (i.e. non-renormalisable) operators in the SM fields. Assuming that the leading new-physics effects arise at dimension six, this Lagrangian reads

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_j \frac{C_j}{\Lambda^2} \mathcal{O}_j + \mathcal{O}(1/\Lambda^3), \qquad (2)$$

where $\mathcal{L}_{\rm SM}$ is the SM Lagrangian, and the Wilson coefficients C_j encode the BSM details of the theory. Among the 3045 free parameters in this general SMEFT Lagrangian of Eq. (2) [37,38], only a few are relevant for top quark physics.

We consider a scenario in which CP is conserved, and we next assume that new physics only couples to the weak doublet of left-handed top and bottom quarks (Q) and the right-handed weak singlets (t and b) of third-generation quarks (as well as to SM bosons). Moreover, bosonic operators leading to flavour-universal effects are discarded, we approximate the CKM matrix by the identity matrix, and all Yukawa couplings but those of the top and bottom quarks are neglected. In order to further reduce the number of free parameters, we consider a $U(2)_q \times U(2)_u \times U(2)_d$ flavour symmetry

among the quarks of the first and second generations, in agreement with the principle of minimal flavour violation [60–62]. Differences between the first and second-generation quarks are thus ignored, and we subsequently introduce the generic notation q for a left-handed weak doublet of first-generation or second-generation quark fields, and u and d for the corresponding right-handed weak singlets of up-type and down-type quark fields.

In our analysis, we aim to leverage the detector-simulation capabilities of the MADANALYSIS 5 and RIVET frameworks (including our implementation of HEPTOPTAGGER) to realistically explore the effects of effective operators on the reconstruction performance of boosted top quarks. Among the full set of potentially impactful SMEFT operators [39], only eight of them are not too strongly constrained by other means [40–45], so that an investigation of pair-production and decay of boosted top quarks could offer new handles on them. They read, in the notation of Ref. [42],

$$\begin{aligned}
\mathcal{O}_{Qq}^{1,8} &= (\bar{Q}\gamma_\mu T^A Q) \ (\bar{q}\gamma^\mu T^A q), \\
\mathcal{O}_{Qq}^{3,8} &= (\bar{Q}\gamma_\mu T^A \sigma^I Q) \ (\bar{q}\gamma^\mu T^A \sigma^I q), \\
\mathcal{O}_{Qq}^{3,1} &= (\bar{Q}\gamma_\mu \sigma^I Q) \ (\bar{q}\gamma^\mu \sigma^I q), \\
\mathcal{O}_{tu}^{8} &= (\bar{t}\gamma_\mu T^A t) \ (\bar{u}\gamma^\mu T^A u), \\
\mathcal{O}_{td}^{8} &= (\bar{t}\gamma_\mu T^A t) \ (\bar{d}\gamma^\mu T^A d), \\
\mathcal{O}_{Qu}^{8} &= (\bar{Q}\gamma_\mu T^A Q) \ (\bar{u}\gamma^\mu T^A u), \\
\mathcal{O}_{Qd}^{8} &= (\bar{Q}\gamma_\mu T^A Q) \ (\bar{d}\gamma^\mu T^A d), \\
\mathcal{O}_{tq}^{8} &= (\bar{q}\gamma_\mu T^A q) \ (\bar{t}\gamma^\mu T^A t), \qquad (3)
\end{aligned}$$

where the matrices T^A stand for the generators of $SU(3)_c$ in the fundamental representation, and the matrices σ^I are the usual Pauli matrices.

3.2 Top tagging performance in the presence of non-vanishing SMEFT operators

In order to assess how non-zero values for the Wilson coefficients associated with the SMEFT operators of Eq. (3) affect top quark tagging performance, we make use of MADGRAPH5_AMC@NLO version 3.0.3 [63] to generate parton-level events describing top-antitop production and their semi-leptonic decay at the LHC (operating at a centre-of-mass energy of 13 TeV). We rely on leading-order matrix elements convolved with the leading-order set of NNPDF3.0 parton distribution functions [64] provided through the LHAPDF6 library [65]. For efficiency reasons, the Monte Carlo event generation was kinematically biased to high scales, and we required that the invariant mass of the produced $t\bar{t}$ system satisfies $m_{t\bar{t}}^{\rm truth} > 950$ GeV. These fixed-order events are matched with parton showering and hadronisation as modelled by PYTHIA version 8.2 [66]. Background events are generated with the same toolchain, but by consider-

ing the production of a leptonically-decaying W boson in association with a pair of b jets (and two additional jets), $pp \to W b\bar{b} +$ jets.

Our canonical analysis was implemented in RIVET version 3 [33].[4] It employs FASTJET version 3.3.3 [57] for event reconstruction, and HEPTOPTAGGER version 2 [10] in its default configuration. We remind that the latter has been tuned on boosted top quarks with properties as expected from their SM production and decay, which may thus not be the best for scenarios in which SMEFT effects change the properties of the produced tops. In our usage of HEPTOPTAGGER, we turn on the 'optimal R' option. This allows the tagging algorithm to determine the minimum choice for the fat jet reconstruction radius to ensure that the reconstructed top jet includes a three-prong structure (as expected from standard top quark decays).

Our event reconstruction is achieved by first defining a collection of 'small jets' through the clustering of all visible hadron-level final-state objects with a pseudo-rapidity $|\eta| < 4.5$, muons excepted. We use the anti-k_T jet algorithm [58] with radius parameter $R = 0.4$, and then impose a minimum transverse-momentum requirement of $p_T > 30$ GeV on the reconstructed small jets. Next, we define a collection of 'fat jets' from the same hadron-level objects. This collection is constructed by using the Cambridge-Aachen algorithm [47–49] with a radius parameter $R = 1.5$. We impose a minimum transverse momentum requirement of $p_T > 200$ GeV on the reconstructed fat jets.

Lepton candidates (i.e. electrons and muons) are required to satisfy basic momentum and pseudo-rapidity criteria, $p_T > 10$ GeV and $|\eta| < 2.5$. At this stage, ΔR-based isolation is enforced in order to remove the overlap between the lepton collection and the two jet collections. We remove from the small-jet collection any small jet j lying in the vicinity of a lepton ℓ by an angular distance $\Delta R(\ell, j) < 0.1$, and we then discard any lepton lying at a distance $\Delta R(\ell, j) < 0.4$ of any of the remaining small jets. Moreover, we define b jets as small jets with $p_T > 30$ GeV and with a ghost-associated b-hadron with $p_T > 5$ GeV [67,68].

After reconstruction, we select events whose topology is compatible with that expected from the production of a pair of boosted top quarks that decays semi-leptonically. We require that each selected event features one lepton with at least 50 GeV of transverse momentum, a minimum missing transverse energy $\not{E}_T > 30$ GeV, as well as at least two small b jets and two small light jets. Next, we reconstruct the leptonically-decaying W boson that we consider on-shell. This assumption implies that the invariant mass of the system comprising the lepton and the missing momentum is equal to the mass m_W of the W boson, which allows us to determine

---

[4] An equivalent implementation in MADANALYSIS 5 produced similar results.

the longitudinal component $\not{p}_z$ of the missing momentum,

$$
\begin{aligned}
m_W^2 &= (\mathbf{p}_\ell + \not{\mathbf{p}})^2 \\
&= \left(E_\ell + \sqrt{\not{p}_T^2 + \not{p}_z^2} \right)^2 - \sum_i \left(p_{\ell,i} + \not{p}_i \right)^2 \\
&\quad \text{with } i \in \{x, \ y, \ z\}.
\end{aligned}
\tag{4}
$$

In the above expression, $\mathbf{p}_\ell = (p_{\ell,x}, p_{\ell,y}, p_{\ell,z})$ denotes the three-momentum of the lepton, $\not{\mathbf{p}} = (\not{p}_x, \not{p}_y, \not{p}_z)$ is the missing three-momentum, and E_ℓ stands for the energy of the lepton. From the solution to Eq. (4), we can define the four-momentum of the leptonically-decaying W boson W_L^{rec}. In the case where this equation has two solutions, we arbitrarily choose the smallest value for $\not{p}_z$. Moreover, when it has no solution, we set the associated discriminant to 0 and use the resulting solution.

In order to reconstruct the leptonically-decaying top quark, we match this reconstructed W boson with one of the b jets by minimising the difference between the top mass m_t and the invariant mass $m[W_L^{\text{rec}} \oplus b]$ of the system constituted of the reconstructed W boson W_L^{rec} and the b jet. This is achieved through a $\Delta\chi^2$ minimisation,

$$
\Delta\chi^2 = \frac{|m_t - m[W_L^{\text{rec}} \oplus b]|^2}{\sigma^2} \equiv \frac{|m_t - m(t_L^{\text{rec}})|^2}{\sigma^2},
\tag{5}
$$

with a mass-resolution parameter $\sigma = 40$ GeV. The b-jet matched in this leptonic-top reconstruction is denoted by b_L in the following text.

Figure 1 illustrates the features of the reconstruction of the leptonic branch of the process. It shows the distribution in the invariant mass $m(W_L^{\text{rec}})$ of the reconstructed W boson (upper panel) and that in the invariant mass $m(t_L^{\text{rec}})$ of the reconstructed top quark (lower panel). Predictions are displayed both for the $t\bar{t}$ signal (red) and the associated background (blue). These results demonstrate that most signal events exhibit an on-shell leptonically-decaying W-boson and an on-shell associated top quark. However, the tails of the distributions extend quite significantly away from the peak values for the two spectra. This originates from the inefficiencies inherent to the kinematic fit performed in Eq. (4), which could lead to zero, one, or two solutions for $\not{p}_z$. Consequently, the reconstructed mass of the W_L^{rec} boson (upper panel of Fig. 1) exhibits a plateau at values lower than the true W mass. This impacted our choice for the numerical value of the resolution parameter used in the χ^2 fit of Eq. (5), which then leads to a quite broad peak around the true top mass for the distribution in the reconstructed top mass (lower panel of Fig. 1).

In the next step of our analysis, we study to which extent a hadronically-decaying top quark can be reconstructed from the event's final state. We start from the fat-jet collection and discard any fat jet J that lies at
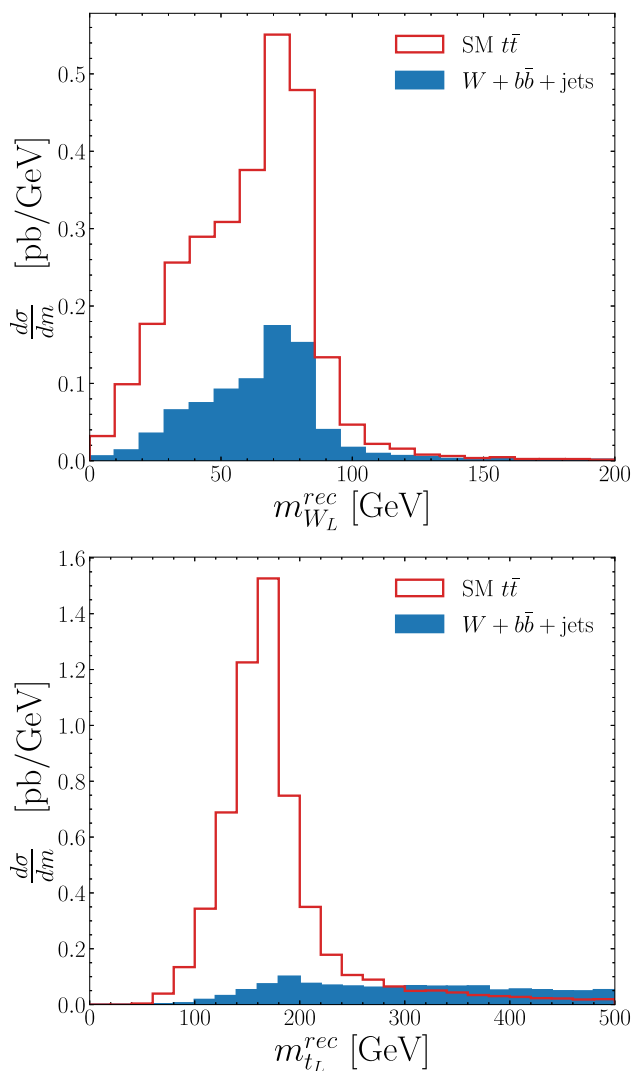
Fig. 1 Invariant mass spectra relevant to the reconstruction of the leptonically decaying top quark. We display the invariant mass $m(W_{\mathrm{L}}^{\mathrm{rec}})$ of the reconstructed W boson (upper panel), as well as that ($m(t_{\mathrm{L}}^{\mathrm{rec}})$) of the reconstructed top quark (lower panel). Predictions are shown for both the $t\bar{t}$ signal (red) and the associated background (blue)

angular distance $\Delta R(J, t_{\mathrm{L}}^{\mathrm{rec}}) \leq 1.5$ of the reconstructed leptonically-decaying top quark $t_{\mathrm{L}}^{\mathrm{rec}}$. Next, we discard all fat jets found near the b_{L} jet, i.e. lying within a angular distance $\Delta R(J, b_{\mathrm{L}}) \leq 1.5$. Finally, we reject events that do not comprise at least one fat jet that includes a (small) b-jet. This condition is implemented by requiring that there is a fat jet J such that a b-jet different from the b_{L} jet lies at a distance $\Delta R(J, b) < 1.5$ from it. We then test whether the hardest of the remaining fat jet is top-tagged by HEPTOPTAGGER.

We now introduce a few useful quantities in order to assess the performance of HEPTOPTAGGER. First, we classify a truth-level top quark as "on-shell" when its invariant mass is in the range $[m_t - 15 \text{ GeV}, m_t + 15 \text{ GeV}]$, and define the quantity $T_{t\bar{t}}$ as the number of $t\bar{t}$ events featuring two such on-shell top quarks. Next, we denote by C_{t_H} the number of

events for which the reconstructed hadronic top quark lies within an angular distance $\Delta R < 1.2$ from the corresponding truth-level object when the latter is on-shell.[5] Similarly, $C_{t_{\mathrm{L}}^{\mathrm{rec}}}$ stands for the number of events for which the reconstructed leptonically-decaying top quark lies at a distance $\Delta R < 1.2$ of its truth-level counterpart when it is on-shell. The quantities C_{t_H} and $C_{t_{\mathrm{L}}^{\mathrm{rec}}}$ hence refer to the number of events for which the reconstructed top quarks are matched with the corresponding truth-level objects so that reconstruction is deemed correct.

With the first set of three coloured columns displayed on the left of Fig. 2, we show the resulting reconstruction efficiency defined as the ratio of the number of events featuring correctly reconstructed hadronic and leptonic top quarks to the number of events including two truth-level on-shell top quarks, i.e. the self-explanatory quantity

$$\varepsilon = \frac{C_{t_H} \wedge C_{t_{\mathrm{L}}^{\mathrm{rec}}}}{T_{t\bar{t}}} \equiv \frac{\left(C_{t_H} \text{ AND } C_{t_{\mathrm{L}}^{\mathrm{rec}}} \right)}{T_{t\bar{t}}}. \qquad (6)$$

This efficiency is given when the baseline cuts described above are imposed (red), when an additional selection of $m_{t\bar{t}}^{\mathrm{truth}} > 1$ TeV is enforced (blue), and finally, when we require $m_{t\bar{t}}^{\mathrm{truth}} > 1.5$ TeV (green). The error bars represent the related Monte Carlo statistical uncertainty. We observe that about 50% of the SM events with on-shell $t\bar{t}$ production are correctly reconstructed, this number slightly increasing when we focus more deeply on the boosted regime (i.e. with a larger $m_{t\bar{t}}^{\mathrm{truth}}$ cut).

The efficiency, however, increases once one of the SMEFT operators of Eq. (3) is turned on, as shown in the rest of Fig. 2 (the dashed lines being guidelines for the comparison with the case of the SM). Here, the signal is simulated by implementing the Lagrangian and operators of Eqs. (2) and (3) in FEYNRULES as specified in Refs. [69,70]. This is then used to generate a UFO [71] model to be used within MADGRAPH5_AMC@NLO so that events could be generated through the same toolchain as that described at the beginning of this section. However, whereas we include the interference of dimension-six contributions with SM diagrams, squared SMEFT contributions (thus formally of dimension-eight) are truncated away. The increase in efficiency observed in Fig. 2 can be traced back not only to a slight increase in the signal cross section, but also to a change in the event topology enhancing HEPTOPTAGGER's ability to correctly tag the boosted, hadronically-decaying top quark. To prove this statement, we display in Fig. 3 the efficiency ε' of correctly tagging the leptonic top $t_{\mathrm{L}}^{\mathrm{rec}}$ regardless of the hadronic branch of the events,

$$\varepsilon' = \frac{C_{t_{\mathrm{L}}^{\mathrm{rec}}}}{T_{t\bar{t}}}. \qquad (7)$$

---

[5] In our notation, T is related to 'true' and C to 'corresponding'.
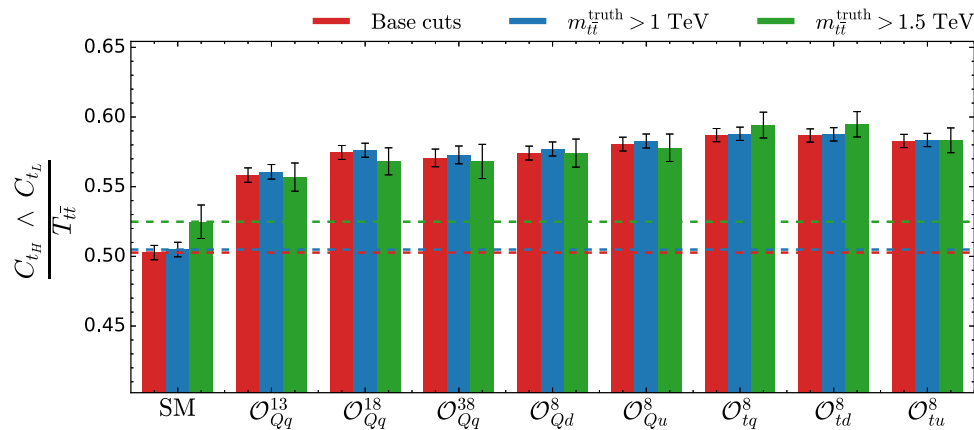
**Fig. 2** Efficiency associated with the reconstruction of one leptonic and one hadronic top quark, estimated relatively to the number of events containing two on-shell top quarks. Results are shown after the analysis baseline cuts (red), an additional $m_{t\bar{t}}^{\text{truth}} > 1$ TeV cut (blue), and an extra $m_{t\bar{t}}^{\text{truth}} > 1.5$ TeV cut (green). We consider the case of the SM (first column), as well as when eight different SMEFT operators are turned on (next columns)
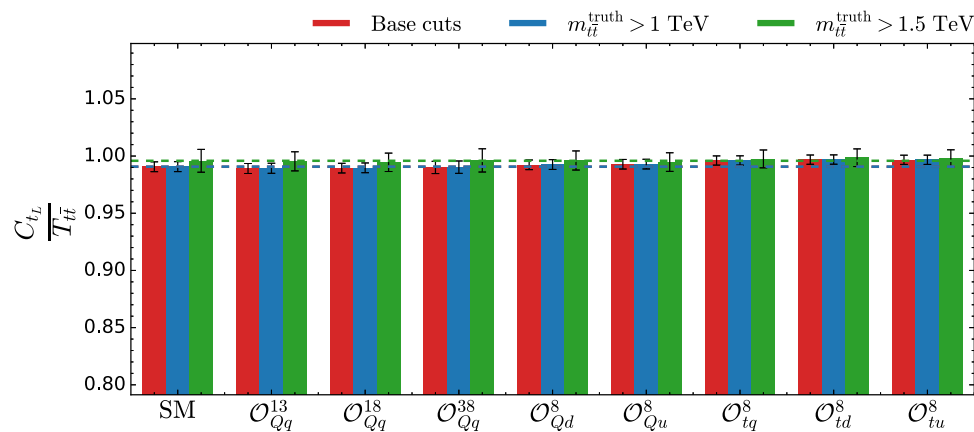
Fig. 3 Same as in Fig. 2 but for the efficiency associated with the reconstruction of one leptonic top quark, estimated relatively to the number of events containing two on-shell top quarks

As can be seen in this figure, the efficiency ε' is almost 100% for all considered scenarios (both in terms of new-physics setup and the parton-level $m_{t\bar{t}}^{\text{truth}}$ cut). This confirms that the global suppression of the efficiency ε shown in Fig. 2 (relative to ε') originates solely from the tagging of the hadronic top quark, and is therefore related to the performance of HEP-TOPTAGGER. The latter can thus directly be assessed from the quantity ε, and is different between SM $t\bar{t}$ events and those including the interference of top-related SMEFT operators with the SM.

Our results demonstrate that the performance of HEP-TOPTAGGER could be strongly impacted by the physics model that is used as a reference during its tuning. Including effective operators such as those in Eq. (3) favours the production of a boosted top-antitop pair more than in the SM, as expected from operators sensitive to the event's energy scale. While in this case the presence of operators not included in the HEPTOPTAGGER tuning enhances the reconstruction efficiency, this is not generally true, and a tuning based on potential EFT contributions could find different optimal tagging parameters.

Importantly, analyses assuming SM-like HEPTOPTAG-GER reconstruction efficiencies, would underestimate the reconstruction and tagging efficiency for any data $t\bar{t}$ events involving these operators, and would hence systematically overestimate the magnitude of the corresponding Wilson coefficient. This observation reinforces the importance of using operator-dependent reconstruction efficiencies in SMEFT fits to boosted top quark data.

The presented efficiencies are, however, normalised to the number of events featuring an on-shell $t\bar{t}$ pair. The obtained increase in the tagging efficiency ε in the presence of SMEFT operators may, therefore, also be related to a different probability of getting at least one off-shell top quark in the events.
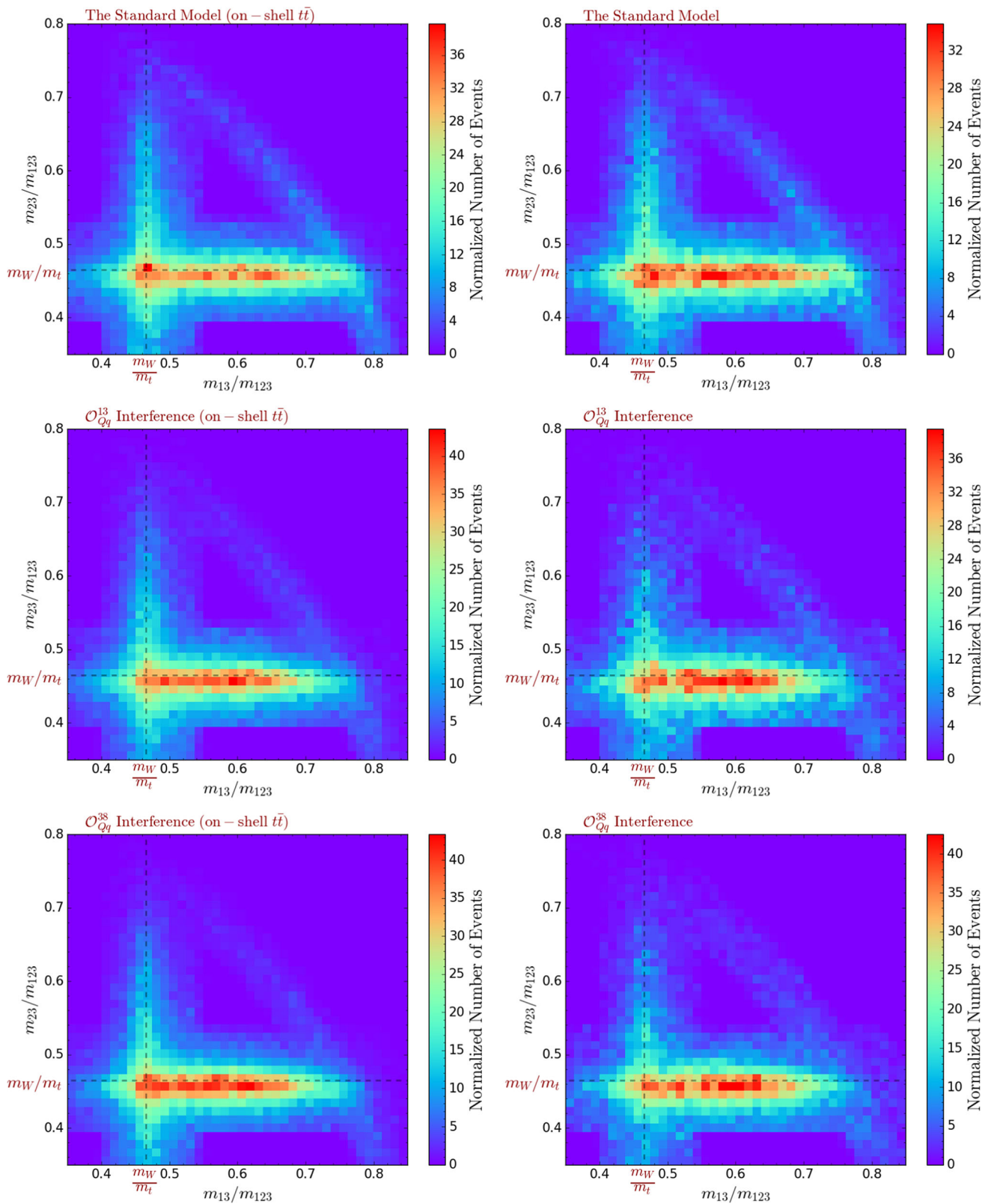
Fig. 4 Dalitz plots depicting the invariant mass ratios m_{13}/m_{123} and m_{23}/m_{123} where the indices refer to a specific jet among those comprising the reconstructed hadronically-decaying top quark. We show predictions when the on-shellness of the top quark is enforced (left column) and when there is no restriction on the invariant mass of the top quarks at parton level (right column). We consider the case of the SM (top row) and that of scenarios with one SMEFT operator turned on, namely $\mathcal{O}_{Qq}^{3,1}$ (middle row) and $\mathcal{O}_{Qq}^{3,8}$ (bottom row)

This problem is addressed by the Dalitz-plot heat-maps shown in Fig. 4, which depict the on-shellness of the produced hadronic top quark. In these figures, we display the correlations between two ratios of invariant masses, m_{13}/m_{123} and m_{23}/m_{123}. The three integers 1, 2 and 3 denote the three (p_{T}-ordered) subjets comprised in the hadronically-decaying boosted top quark, so that m_{123} stands for the invariant mass of the three-subjet system, m_{13} for the invariant mass the system made of the leading and third subjets, and m_{23} for that of the system made of the second and third subjets. We present results by restricting the events to those events featuring on-shell top quarks (left column) and for the entire generated samples (right column). Moreover, we explore the difference between the SM (top row), a scenario in which the $\mathcal{O}_{Qq}^{3,1}$ operator of Eq. (3) is turned on (middle row), and a scenario in which the $\mathcal{O}_{Qq}^{3,8}$ operator of Eq. (3) is turned on (bottom row).

As can be seen, the jet combinatorics are correctly resolved in most events in the case of the SM. The leading jet is most often that originating from the two-body $t \to Wb$ decay (with the b-tagging information being ignored), and the next two jets are those stemming from the hadronic W-boson decay. The distribution of the m_{23}/m_{123} ratio is indeed concentrated around m_W/m_t for the two subfigures of the upper row of Fig. 4. The spread around this value is more pronounced when no restriction is enforced on the invariant mass of the top quarks at parton level, as observed from a comparison of the predictions shown in the top-left and top-right figures. This can be easily explained by the inefficiency of HEPTOPTAGGER to correctly tag off-shell top jets, as, by default, the algorithm has been tuned on events featuring on-shell top quarks.

This situation changes slightly when EFT operators are enabled (middle and lower rows of Fig. 4). First, although the associated amplitude does not feature any intermediate W boson (as the decay of the top quark proceeds via a single four-fermion operator), the interference with the SM diagrams (our predictions being truncated at dimension-six) is sufficient to keep the properties that the leading jet is the b-jet, and that the next two jets can be paired to reconstruct a hadronically-decaying W boson. It is additionally noticeable that the effective operators considered affect the reconstructed top quark so that the latter is naturally more often on-shell (and more boosted due to the energy growth inherent to the effective-theory paradigm). Consequently, we can expect better performance of HEPTOPTAGGER, which confirms what was already found in Fig. 2.

3.3 Boosted tops as a probe to new physics in the SMEFT

In this section, we explore how the findings of Sect. 3.2 affect the sensitivity of the LHC to SMEFT effects originat-

Table 3 Number of $t\bar{t}$ and $Wb\bar{b}$+jets SM events surviving each step of our analysis, presented together with their respective selection efficiency ε. The results are normalised to an integrated luminosity of 300 fb^{-1}. In the last row of the table, we provide two alternative means to assess the analysis significance, namely the S/B and $S/\sqrt{B}$ ratios where S and B are the number of $t\bar{t}$ and background events passing all cuts

Selections		Background		SM signal	
		Events	ε	Events	ε
1.	Initial	124,264.50	–	346,061.01	–
2.	$N_\ell = 1$	85,506.63	0.688	227,116.93	0.656
3.	$p_{\mathrm{T}}(\ell_1) > 30$ GeV	46,668.91	0.546	133,628.45	0.588
4.	$N_b \geq 2$	23,515.04	0.504	85,299.59	0.638
5.	$N_{\mathrm{light\ jets}} > 2$	21,126.59	0.898	75,787.27	0.888
6.	$\not{E}_{\mathrm{T}} > 30$ GeV	18,002.61	0.852	66,834.09	0.882
7.	$N_{\mathrm{fat\text{-}jet}} > 0$	5231.95	0.291	53,417.98	0.799
8.	$N_{t_H^{\mathrm{rec}}} > 0$	274.58	0.052	24,478.32	0.458
9.	$m_{t\bar{t}}^{\mathrm{rec}} > 950$ GeV	164.26	0.598	23,260.59	0.950
	S/B			141.61	
	$S/\sqrt{B}$			1814.91	

ing from the operators of Eq. (3). We begin by providing, in Table 3, the numbers of events surviving each of the selection cuts introduced in the previous section, both for the $t\bar{t}$ signal and the $Wb\bar{b}$ + jets background. Our results are normalised to an integrated luminosity of 300 fb^{-1}, and we additionally estimate the efficiencies associated with each cut, which we define as the ratio of the number of events surviving a given cut to the number of events surviving the previous cut. Whereas the last cut on the invariant mass of the reconstructed $t\bar{t}$ system (i.e. the ninth one in the table, $m_{t\bar{t}}^{\mathrm{rec}} > 950$ GeV) is not necessary for physics-analysis purposes, it is required to match the Monte Carlo signal-generation cut implemented in Sect. 3.2 (to enable a more efficient event-generation process in the boosted regime).

As already noticeable from the results introduced earlier in this manuscript, for instance from the invariant-mass spectra displayed in Fig. 1, the events surviving the entire selection are primarily dominated by signal events, which hence have large expected event-counts. This is further reflected in the S/B and $S/\sqrt{B}$ ratios provided as significance estimators in the lower rows of Table 3, these two metrics being evaluated in terms of the number of signal events S and background events B passing all the analysis cuts. The background is thus fully under control in our study, so a shape analysis can be implemented to study how kinematic distributions can be best used to constrain the SMEFT-operators' Wilson coefficients.

To do this, we first increase the final selection cut to maximise sensitivity by probing more deeply boosted top-antitop production. In the following, we hence consider either $m_{t\bar{t}}^{\mathrm{rec}} > 1$ TeV or $m_{t\bar{t}}^{\mathrm{rec}} > 1.5$ TeV. The sensitivity of the LHC to a given SMEFT operator is derived through the evaluation

of a χ^2 test-statistic in an asymptotic scheme that involves deviations of SMEFT predictions relative to the associated SM predictions for a given set of observables. Our analysis explores simultaneously the distributions of the following observables:

- the invariant mass $m_{t\bar{t}}^{\text{rec}}$ of the di-top system;
- the transverse momentum $p_{\text{T}}(j^{R=1.5})$ of the leading fat-jet;
- the transverse momenta $p_{\text{T}}(j_1^{R=0.4})$, $p_{\text{T}}(j_2^{R=0.4})$ and $p_{\text{T}}(j_3^{R=0.4})$ of the three leading small-R jets;
- the transverse-momentum spectrum $p_{\text{T}}(t_{\text{H}})$ of the reconstructed hadronic top quark;
- the transverse-momentum spectrum $p_{\text{T}}(t_{\text{L}}^{\text{rec}})$ of the reconstructed leptonic top quark;
- the rapidity difference $\Delta y(t_{\text{L}}^{\text{rec}}, t_{\text{H}})$ between the two reconstructed top quarks;
- and the azimuthal-angle difference $\Delta\varphi(t_{\text{L}}^{\text{rec}}, t_{\text{H}})$ between the two reconstructed top quarks.

In order to estimate the χ^2 value associated with a specific SMEFT scenario, each of the nine histograms considered was divided into 25 bins (20 and 16 for the $\Delta y(t_{\text{L}}^{\text{rec}}, t_{\text{H}})$ and $\Delta\varphi(t_{\text{L}}^{\text{rec}}, t_{\text{H}})$ distributions respectively), and we calculated the quantity

$$\chi^2 = \sum_i \frac{\left(N_i^{\text{exp}} - N_i^{\text{obs}}\right)^2}{\sqrt{N_i^{\text{obs}} + \left(\Delta_{\text{sys}} N_i^{\text{obs}}\right)^2}}, \tag{8}$$

in which we sum over all bins and all histograms. The SM predictions are taken as the null hypothesis, N_i^{exp} denoting hence the expected number of events in the SM for a given observable and bin i, N_i^{obs} standing for the corresponding SMEFT predictions, and $\Delta_{\text{sys}} N_i^{\text{obs}}$ referring to the error on the SMEFT predictions. In other words, we enforce that the pseudo-data corresponding to the SM scenario (i.e. the origin of the Wilson coefficient parameter space) corresponds to the background expectation with suppressed statistical and systematical fluctuations, which consists, therefore, of an Asimov dataset. The above χ^2 test is thus asymptotically equivalent to a profile likelihood ratio $\Delta\chi^2 = \chi^2_{\text{SMEFT}} - \chi^2_{\text{best}}$ for a given SMEFT scenario with an implicit best-fit reference model evaluated in the case of the SM (therefore with $\chi^2_{\text{best}} = 0$). Without explicitly performing any profiling, we thus estimate the sensitivity of a profile-likelihood fit by comparing the obtained χ^2 values with that expected from a χ^2 distribution with one degree of freedom. In practice, however, profiled constraints could be slightly weaker due to a less perfect fit of observed data to the background model.

In Table 4, we provide information on the observable found to provide the strongest sensitivity to each SMEFT operator. The results are shown for the two cuts on the invariant mass considered, $m_{t\bar{t}}^{\text{rec}} > 1$ TeV (upper panel of

Table 4 Observable driving the sensitivity of the LHC (at 68% confidence level) to a given SMEFT operator from Eq. (3) (first column). We consider both a perfect situation without systematics ($\Delta_{\text{sys}} = 0$, second and fourth columns), and one with 10% of systematics ($\Delta_{\text{sys}} = 10\%$, third and fifth columns). Moreover, we present results for 300 fb^{-1} and 3 ab^{-1}, and for an invariant mass cut of $m_{t\bar{t}}^{\text{rec}} > 1$ TeV (upper panel) and $m_{t\bar{t}}^{\text{rec}} > 1.5$ TeV (lower panel)

Mass req.	$m_{t\bar{t}}^{\text{rec}} > 1$ TeV			
Lumi. [ab^{-1}]	0.3		3	
Δ_{sys} [%]	0	10	0	10
$\mathcal{O}_{Qq}^{1,8}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{Qq}^{3,8}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{Qq}^{3,1}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{tu}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$
$\mathcal{O}_{td}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$
$\mathcal{O}_{Qu}^{8}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{Qd}^{8}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{tq}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$
Mass req.	$m_{t\bar{t}}^{\text{rec}} > 1.5$ TeV			
Lumi. [ab^{-1}]	0.3		3	
Δ_{sys} [%]	0	10	0	10
$\mathcal{O}_{Qq}^{1,8}$	$p_{\text{T}}(t_{\text{H}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{H}})$	$\Delta y_{t\bar{t}}$
$\mathcal{O}_{Qq}^{3,8}$	$p_{\text{T}}(t_{\text{H}})$	$p_{\text{T}}(j_1^{R=1.5})$	$p_{\text{T}}(t_{\text{H}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{Qq}^{3,1}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$\Delta y_{t\bar{t}}$
$\mathcal{O}_{tu}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$
$\mathcal{O}_{td}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$
$\mathcal{O}_{Qu}^{8}$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$
$\mathcal{O}_{Qd}^{8}$	$p_{\text{T}}(t_{\text{H}})$	$p_{\text{T}}(t_{\text{L}}^{\text{rec}})$	$p_{\text{T}}(t_{\text{H}})$	$\Delta y_{t\bar{t}}$
$\mathcal{O}_{tq}^{8}$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$	$p_{\text{T}}(\ell_1)$

Table 4) and $m_{t\bar{t}}^{\text{rec}} > 1.5$ TeV (lower panel of Table 4). Moreover, we consider LHC luminosities of 300 fb^{-1} and 3000 fb^{-1}, and two different options for the amount of systematics Δ_{sys} used in Eq. (8). We take as a reference the ideal situation in which there are no systematic uncertainties ($\Delta_{\text{sys}} = 0$), as well as a more realistic situation in which we set $\Delta_{\text{sys}} = 10\%$. In our procedure to extract this information, we define the sensitivity on the basis of a 68% confidence level. When we consider a moderate definition of the boosted regime with $m_{t\bar{t}}^{\text{rec}} > 1$ TeV, the sensitivity is always driven by the distribution in the transverse momentum of either the leptonically-decaying top quark ($p_{\text{T}}(t_{\text{L}}^{\text{rec}})$) or of the lepton originating from the decay of this top quark ($p_{\text{T}}(\ell_1)$). The information brought by the hadronic branch of the event is found to be sub-leading for all SMEFT operators and systematic-uncertainty assumptions. However, the situation changes when the boosted regime is probed more

deeply through the tighter cut $m_{t\bar{t}}^{\text{rec}} > 1.5\,\text{TeV}$. Here, both top quarks are reconstructed and tagged more accurately (in particular through the better performance of HEPTOPTAGGER in a SMEFT scenario, see Sect. 3.2). This leads to an increased discovery potential through use of a larger set of contributing observables. This statement is illustrated in the lower panel of the table, which displays a greater variability in the leading observable driving the sensitivity of the LHC to a given SMEFT operator, with the $\mathcal{O}_{Qq}^{1,8}$, $\mathcal{O}_{Qq}^{3,8}$, $\mathcal{O}_{Qq}^{3,1}$, and $\mathcal{O}_{Qd}^{8}$ operators now most sensitive to either hadronic-top or $t\bar{t}$-system observables.

Our final projections of SMEFT Wilson-coefficient expected limits, assuming the SM, are shown in Fig. 5. We derive the sensitivity of the LHC to each of the operators considered, making use of the procedure described above. We present bounds on the associated Wilson coefficients, both for an integrated luminosity of 300 fb^{-1} (blue) and 3000 fb^{-1} (red), and for the two options explored for the level of systematics, namely $\Delta_{\text{sys}} = 0$ (shaded bars) and 10% (solid bars). In addition, we distinguish the case in which we pre-select at parton-level on-shell $t\bar{t}$ events (left subfigures) and that in which we analyse the full event sample generated (right subfigures). As for the previous discussion, we first implement a relatively inclusive requirement of 1 TeV on the invariant mass of the reconstructed $t\bar{t}$ system (upper row) and as well as a more stringent $m_{t\bar{t}}^{\text{rec}} > 1.5\,\text{TeV}$ cut (bottom row).

We find limits on $|C/\Lambda|$ that lie in the 0.1–$1\,\text{TeV}^{-1}$ range. This means that for Wilson coefficients satisfying $C \sim 1$, effective scales in the 1–5 TeV range can be probed. Conversely, for TeV-scale new physics, couplings of $\mathcal{O}(0.1)$ can be reached. The bounds are found to be mildly more constraining with the increase in luminosity as well as with a harder cut on $m_{t\bar{t}}^{\text{rec}}$, as expected, and the impact of off-shell top-antitop production is additionally found to be sub-leading. Such a sensitivity is of comparable size with that estimated on the basis of global fits (see e.g. predictions from Ref. [42]), which demonstrates the potential of including dedicated analyses of boosted top quark pair production and decay in SMEFT global fits. Global fits of LHC Run 2 data indeed indicate that $|C/\Lambda|$ has to be smaller than about 0.1–$1\,\text{TeV}^{-1}$ too. Our results should however additionally be compared with individual limits extracted from fits of a large set of observables when one SMEFT operator is considered at a time (for a fairer comparison). Such fits lead to bounds on $|C/\Lambda|$ of $\mathcal{O}(0.1)\,\text{TeV}^{-1}$ [44], which are thus comparable with the findings of Fig. 5. Whereas exploiting boosted top quark production is already known to have a strong constraining power on individual operators (for instance in the context of top dipole moments, where it has been shown to significantly improve the bounds by a factor of a few [72]), a detailed quantitative analysis of its impact lies beyond the scope of this paper. Here, we have only investigated how

using a specific boosted-top quark channel could lead to a better assessment of the sensitivity of the LHC to top quark-related SMEFT operators, thanks to a joint usage of a variety of potentially relevant observables and improved top-tagging capabilities in the SMEFT.

4 Conclusion and outlook

Jet substructure methods are known to be among the key players in the search for new phenomena beyond the Standard Model of particle physics. Among these, a set of dedicated techniques are related to the identification of jets originating from the hadronic decay of a boosted top quark. In this paper, we have reported the development of an interface between the HEPTOPTAGGER package and two software tools widely used in the high-energy physics community, namely the MADANALYSIS 5 and RIVET frameworks. Thanks to this development, the many users of these platforms now have the possibility to exploit boosted hadronically-decaying top quarks and their properties in analyses of high-energy physics events for the Large Hadron Collider and beyond.

We have briefly described these two implementations and how to use them. Our developments equip the RIVET toolkit from version 3.1.7, which is available from HEPFORGE (see https://rivet.hepforge.org/), as well as the MADANALYSIS 5 framework from version 2.0.4, available from GITHUB (see https://github.com/MadAnalysis/madanalysis5/releases). Moreover, detailed tutorials exploiting all the possibilities can be found in the ''analyses/examples/EXAMPLE_HTT .cc'' analysis file shipped with RIVET, as well as in the MADANALYSIS 5 tutorial available from https://github.com/MadAnalysis/tutorial_osu.

To illustrate the power of these developments, we have considered the SMEFT framework in which new physics manifests through non-renormalisable operators in the Standard Model fields. We have focused on eight dimension-six, four-fermion operators relevant to the top quark sector, chosen as they are not stringently constrained by current SMEFT global fits. The analysis of the production of pairs of boosted top quarks could therefore provide new handles on associated heavy BSM physics. We have explored this option by first investigating the performance of the HEPTOPTAGGER algorithm in the presence of non-vanishing SMEFT operators. Whereas the algorithm is tuned on SM top-pair production and decay, we have observed that its performance improves further in the presence of the considered additional SMEFT operators in the model's Lagrangian. The energy dependence of the SMEFT operators considered indeed favours the production of very energetic boosted top quarks, with properties enhancing their tagging possibility by the HEPTOPTAGGER method. This observation highlights the importance of
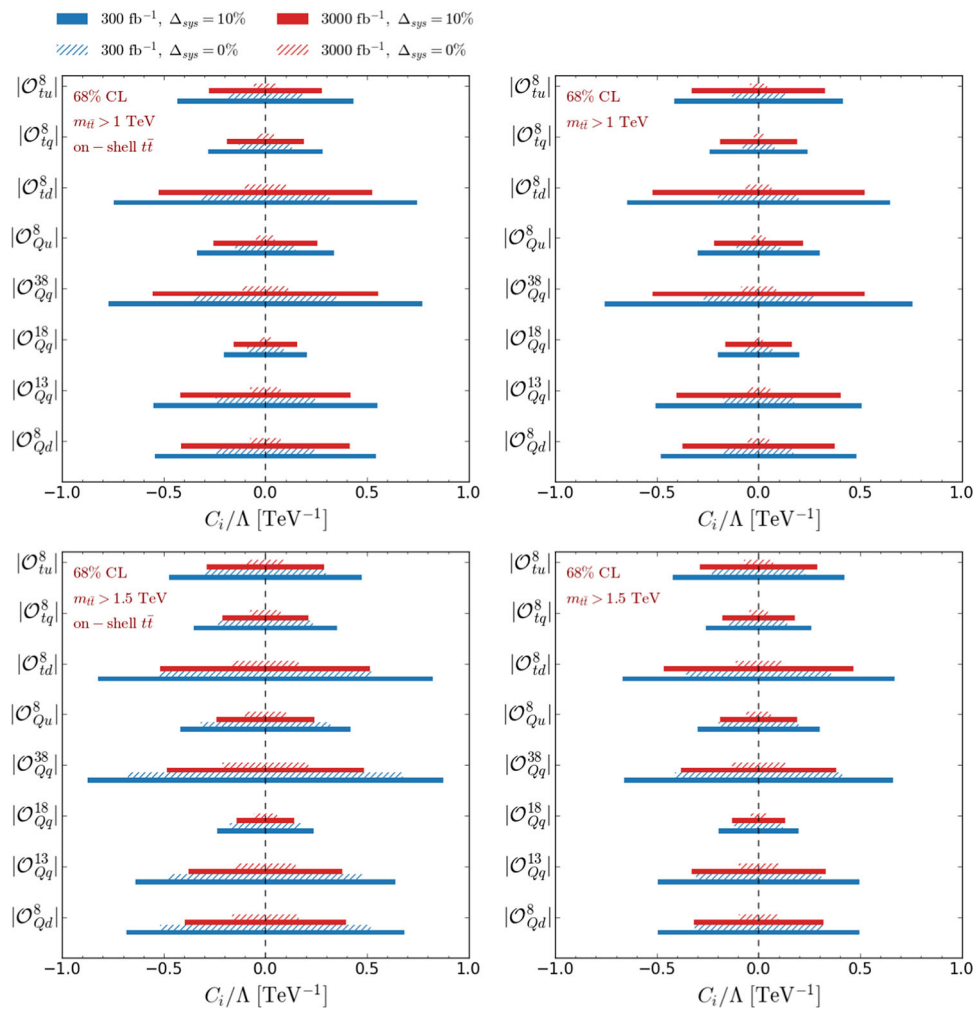
Fig. 5 Sensitivity of the LHC to the various SMEFT operators of Eq. (3). We present predictions for 300 fb^{-1} (blue) and 3000 fb^{-1} (red), $\Delta_{\text{sys}} = 0$ (shaded bars) and 10% (solid bars), and we distinguish an analysis of the full $t\bar{t}$ event sample generated (right column) and after considering new-physics effects upon reconstruction performance when attempting SMEFT parameter fits.

enforcing on-shell top-antitop production (left column). Two analysis cuts on the invariant mass of the reconstructed top pair are imposed, $m_{t\bar{t}}^{\text{rec}} > 1$ TeV (upper panel) and $m_{t\bar{t}}^{\text{rec}} > 1.5$ TeV (lower panel)

Secondly, we have investigated differential observables in boosted top-antitop production following HEPTopTagger tagging, to study how deviations from the Standard Model can best be used to isolate SMEFT effects emerging from the new operators. We have shown that a simple analysis based on HEPTopTagger could lead to bounds comparable with those stemming from other means to constrain SMEFT operators. We hope that this demonstrates the potential of the developments presented in this work and that they will serve the community well in the future.

References

1. M.H. Seymour, Searches for new particles using cone and cluster jet algorithms: a comparative study. Z. Phys. C **62**, 127–138 (1994). https://doi.org/10.1007/BF01559532

2. J.M. Butterworth, A.R. Davison, M. Rubin, G.P. Salam, Jet substructure as a new Higgs search channel at the LHC. Phys. Rev. Lett. **100**, 242001 (2008). https://doi.org/10.1103/PhysRevLett.100.242001. arXiv:0802.2470

3. D.E. Kaplan, K. Rehermann, M.D. Schwartz, B. Tweedie, Top tagging: a method for identifying boosted hadronically decaying top quarks. Phys. Rev. Lett. **101**, 142001 (2008). https://doi.org/10.1103/PhysRevLett.101.142001. arXiv:0806.0848

4. S.D. Ellis, C.K. Vermilion, J.R. Walsh, Techniques for improved heavy particle searches with jet substructure. Phys. Rev. D **80**, 051501 (2009). https://doi.org/10.1103/PhysRevD.80.051501. arXiv:0903.5081

5. A. Abdesselam et al., Boosted objects: a probe of beyond the standard model physics. Eur. Phys. J. C **71**, 1661 (2011). https://doi.org/10.1140/epjc/s10052-011-1661-y. arXiv:1012.5412

6. A.J. Larkoski, I. Moult, B. Nachman, Jet substructure at the large hadron collider: a review of recent advances in theory and machine learning. Phys. Rep. **841**, 1–63 (2020). https://doi.org/10.1016/j.physrep.2019.11.001. arXiv:1709.04464

7. R. Kogler et al., Jet substructure at the large hadron collider: experimental review. Rev. Mod. Phys. **91**, 045003 (2019). https://doi.org/10.1103/RevModPhys.91.045003. arXiv:1803.06991

8. S. Marzani, G. Soyez, M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object phenomenology*, vol. 958 (Springer, Berlin, 2019). https://doi.org/10.1007/978-3-030-15709-8

9. T. Plehn, M. Spannowsky, M. Takeuchi, D. Zerwas, Stop reconstruction with tagged tops. JHEP **10**, 078 (2010). https://doi.org/10.1007/JHEP10(2010)078. arXiv:1006.2833

10. G. Kasieczka, T. Plehn, T. Schell, T. Strebler, G.P. Salam, Resonance searches with an updated top tagger. JHEP **06**, 203 (2015). https://doi.org/10.1007/JHEP06(2015)203. arXiv:1503.05921

11. D.E. Soper, M. Spannowsky, Finding top quarks with shower deconstruction. Phys. Rev. D **87**, 054012 (2013). https://doi.org/10.1103/PhysRevD.87.054012. arXiv:1211.3140

12. A. Butter et al., The machine learning landscape of top taggers. SciPost Phys. **7**, 014 (2019). https://doi.org/10.21468/SciPostPhys.7.1.014. arXiv:1902.09914

13. J. Thaler, K. Van Tilburg, Identifying boosted objects with N-subjettiness. JHEP **03**, 015 (2011). https://doi.org/10.1007/JHEP03(2011)015. arXiv:1011.2268

14. J. Thaler, K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness. JHEP **02**, 093 (2012). https://doi.org/10.1007/JHEP02(2012)093. arXiv:1108.2701

15. J. Thaler, L.-T. Wang, Strategies to identify boosted tops. JHEP **07**, 092 (2008). https://doi.org/10.1088/1126-6708/2008/07/092. arXiv:0806.0023

16. A.J. Larkoski, G.P. Salam, J. Thaler, Energy correlation functions for jet substructure. JHEP **06**, 108 (2013). https://doi.org/10.1007/JHEP06(2013)108. arXiv:1305.0007

17. A.J. Larkoski, I. Moult, D. Neill, Power counting to better jet observables. JHEP **12**, 009 (2014). https://doi.org/10.1007/JHEP12(2014)009. arXiv:1409.6298

18. J. Cogan, M. Kagan, E. Strauss, A. Schwarztman, Jet-images: computer vision inspired techniques for jet tagging. JHEP **02**, 118 (2015). https://doi.org/10.1007/JHEP02(2015)118. arXiv:1407.5675

19. L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman, Jet-images: deep learning edition. JHEP **07**, 069 (2016). https://doi.org/10.1007/JHEP07(2016)069. arXiv:1511.05190

20. P. Baldi, K. Bauer, C. Eng, P. Sadowski, D. Whiteson, Jet substructure classification in high-energy physics with deep neural networks. Phys. Rev. D **93**, 094034 (2016). https://doi.org/10.1103/PhysRevD.93.094034. arXiv:1603.09349

21. L.G. Almeida, M. Backović, M. Cliche, S.J. Lee, M. Perelstein, Playing tag with ANN: boosted top identification with pattern recognition. JHEP **07**, 086 (2015). https://doi.org/10.1007/JHEP07(2015)086. arXiv:1501.05968

22. G. Kasieczka, T. Plehn, M. Russell, T. Schell, Deep-learning top taggers or the end of QCD? JHEP **05**, 006 (2017). https://doi.org/10.1007/JHEP05(2017)006. arXiv:1701.08784

23. S. Macaluso, D. Shih, Pulling out all the tops with computer vision and deep learning. JHEP **10**, 121 (2018). https://doi.org/10.1007/JHEP10(2018)121. arXiv:1803.00107

24. J.Y. Araz, M. Spannowsky, Combine and conquer: event reconstruction with Bayesian ensemble neural networks. JHEP **04**, 296 (2021). https://doi.org/10.1007/JHEP04(2021)296. arXiv:2102.01078

25. S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian et al., An efficient Lorentz equivariant graph neural network for jet tagging. JHEP **07**, 030 (2022). https://doi.org/10.1007/JHEP07(2022)030. arXiv:2201.08187

26. A. Bogatskiy, T. Hoffman, D.W. Miller, J.T. Offermann, PELICAN: permutation equivariant and Lorentz invariant or covariant aggregator network for particle physics. arXiv:2211.00454

27. ATLAS collaboration, M. Aaboud et al., Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC. Eur. Phys. J. C **79**, 375 (2019). https://doi.org/10.1140/epjc/s10052-019-6847-8. arXiv:1808.07858

28. CMS collaboration, A.M. Sirunyan et al., Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. JINST **15**, P06005 (2020). https://doi.org/10.1088/1748-0221/15/06/P06005. arXiv:2004.08262

29. E. Conte, B. Fuks, G. Serret, MadAnalysis 5, a user-friendly framework for collider phenomenology. Comput. Phys. Commun. **184**, 222–256 (2013). https://doi.org/10.1016/j.cpc.2012.09.009. arXiv:1206.1599

30. E. Conte, B. Dumont, B. Fuks, C. Wymant, Designing and recasting LHC analyses with MadAnalysis 5. Eur. Phys. J. C **74**, 3103 (2014). https://doi.org/10.1140/epjc/s10052-014-3103-0. arXiv:1405.3982

31. E. Conte, B. Fuks, Confronting new physics theories to LHC data with MADANALYSIS 5. Int. J. Mod. Phys. A **33**, 1830027 (2018). https://doi.org/10.1142/S0217751X18300272. arXiv:1808.00480

32. A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lonnblad, J. Monk et al., Rivet user manual. Comput. Phys. Commun. **184**, 2803–2819 (2013). https://doi.org/10.1016/j.cpc.2013.05.021. arXiv:1003.0694

33. C. Bierlich et al., Robust independent validation of experiment and theory: Rivet version 3. SciPost Phys. **8**, 026 (2020). https://doi.org/10.21468/SciPostPhys.8.2.026. arXiv:1912.05451

34. S. Weinberg, Phenomenological Lagrangians. Phys. A **96**, 327–340 (1979). https://doi.org/10.1016/0378-4371(79)90223-1

35. C.N. Leung, S.T. Love, S. Rao, Low-energy manifestations of a new interaction scale: operator analysis. Z. Phys. C **31**, 433 (1986). https://doi.org/10.1007/BF01588041

36. W. Buchmuller, D. Wyler, Effective Lagrangian analysis of new interactions and flavor conservation. Nucl. Phys. B **268**, 621–653 (1986). https://doi.org/10.1016/0550-3213(86)90262-2

37. B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek, Dimension-six terms in the standard model Lagrangian. JHEP **10**, 085 (2010). https://doi.org/10.1007/JHEP10(2010)085. arXiv:1008.4884

38. B. Henning, X. Lu, T. Melia, H. Murayama, 2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT. JHEP **08** 016, (2017). https://doi.org/10.1007/JHEP08(2017)016. arXiv:1512.03433

39. D. Barducci et al., Interpreting top-quark LHC measurements in the standard-model effective field theory. arXiv:1802.07237

40. A. Buckley, C. Englert, J. Ferrando, D.J. Miller, L. Moore, M. Russell et al., Constraining top quark effective theory in the LHC Run II era. JHEP **04**, 015 (2016). https://doi.org/10.1007/JHEP04(2016)015. arXiv:1512.03360

41. N.P. Hartland, F. Maltoni, E.R. Nocera, J. Rojo, E. Slade, E. Vryonidou et al., A Monte Carlo global analysis of the standard model effective field theory: the top quark sector. JHEP **04**, 100 (2019). https://doi.org/10.1007/JHEP04(2019)100. arXiv:1901.05965

42. I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou et al., O new physics, where art thou? A global search in the top sector. JHEP **02**, 131 (2020). https://doi.org/10.1007/JHEP02(2020)131. arXiv:1910.03606

43. J. Ellis, M. Madigan, K. Mimasu, V. Sanz, T. You, Top, Higgs, Diboson and electroweak fit to the standard model effective field theory. JHEP **04**, 279 (2021). https://doi.org/10.1007/JHEP04(2021)279. arXiv:2012.02779

44. SMEFiT collaboration, J.J. Ethier, G. Magni, F. Maltoni, L. Mantani, E.R. Nocera, J. Rojo et al., Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC. JHEP **11** 089, (2021). https://doi.org/10.1007/JHEP11(2021)089. arXiv:2105.00006

45. T. Giani, G. Magni, J. Rojo, SMEFiT: a flexible toolbox for global interpretations of particle physics data with effective field theories. arXiv:2302.06660

46. C. Englert, L. Moore, K. Nordström, M. Russell, Giving top quark effective operators a boost. Phys. Lett. B **763**, 9–15 (2016). https://doi.org/10.1016/j.physletb.2016.10.021. arXiv:1607.04304

47. Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, Better jet clustering algorithms. JHEP **08**, 001 (1997). https://doi.org/10.1088/1126-6708/1997/08/001. arXiv:hep-ph/9707323

48. S. Bentvelsen, I. Meyer, The Cambridge jet algorithm: features and applications. Eur. Phys. J. C **4**, 623–629 (1998). https://doi.org/10.1007/s100520050232. arXiv:cond-mat/9803322

49. M. Wobisch, T. Wengler, Hadronization corrections to jet cross-sections in deep inelastic scattering, in Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting), vol. 4, p. 270–279 (1998). arXiv:hep-ph/9907280

50. M. Dasgupta, A. Fregoso, S. Marzani, G.P. Salam, Towards an understanding of jet substructure. JHEP **09**, 029 (2013). https://doi.org/10.1007/JHEP09(2013)029. arXiv:1307.0007

51. D. Krohn, J. Thaler, L.-T. Wang, Jet trimming. JHEP **02**, 084 (2010). https://doi.org/10.1007/JHEP02(2010)084. arXiv:0912.1342

52. S.D. Ellis, C.K. Vermilion, J.R. Walsh, Recombination algorithms and jet substructure: pruning as a tool for heavy particle searches. Phys. Rev. D **81**, 094023 (2010). https://doi.org/10.1103/PhysRevD.81.094023. arXiv:0912.0033

53. Y.-T. Chien, Telescoping jets: probing hadronic event structure with multiple R 's. Phys. Rev. D **90**, 054008 (2014). https://doi.org/10.1103/PhysRevD.90.054008. arXiv:1304.5240

54. S.D. Ellis, A. Hornig, D. Krohn, T.S. Roy, On statistical aspects of qjets. JHEP **01**, 022 (2015). https://doi.org/10.1007/JHEP01(2015)022. arXiv:1409.6785

55. A. Buckley, D. Kar, K. Nordström, Fast simulation of detector effects in Rivet. SciPost Phys. **8**, 025 (2020). https://doi.org/10.21468/SciPostPhys.8.2.025. arXiv:1910.01637

56. J.Y. Araz, B. Fuks, G. Polykratis, Simplified fast detector simulation in MADANALYSIS 5. Eur. Phys. J. C **81**, 329 (2021). https://doi.org/10.1140/epjc/s10052-021-09052-5. arXiv:2006.09387

57. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). https://doi.org/10.1140/epjc/s10052-012-1896-2. arXiv:1111.6097

58. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). https://doi.org/10.1088/1126-6708/2008/04/063. arXiv:0802.1189

59. D. Krohn, J. Thaler, L.-T. Wang, Jets with variable R. JHEP **06**, 059 (2009). https://doi.org/10.1088/1126-6708/2009/06/059. arXiv:0903.0392

60. R.S. Chivukula, H. Georgi, Composite technicolor standard model. Phys. Lett. B **188**, 99–104 (1987). https://doi.org/10.1016/0370-2693(87)90713-1

61. L.J. Hall, L. Randall, Weak scale effective supersymmetry. Phys. Rev. Lett. **65**, 2939–2942 (1990). https://doi.org/10.1103/PhysRevLett.65.2939

62. G. D'Ambrosio, G.F. Giudice, G. Isidori, A. Strumia, Minimal flavor violation: an effective field theory approach. Nucl. Phys. B **645**, 155–187 (2002). https://doi.org/10.1016/S0550-3213(02)00836-2. arXiv:nucl-th/0207036

63. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP **07**, 079 (2014). https://doi.org/10.1007/JHEP07(2014)079. arXiv:1405.0301

64. NNPDF collaboration, R.D. Ball et al., Parton distributions for the LHC Run II. JHEP **04**, 040 (2015). https://doi.org/10.1007/JHEP04(2015)040. arXiv:1410.8849

65. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht et al., LHAPDF6: parton density access in the LHC precision era. Eur. Phys. J. C **75**, 132 (2015). https://doi.org/10.1140/epjc/s10052-015-3318-8. arXiv:1412.7420

66. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015). https://doi.org/10.1016/j.cpc.2015.01.024. arXiv:1410.3012

67. M. Cacciari, G.P. Salam, Pileup subtraction using jet areas. Phys. Lett. B **659**, 119–126 (2008). https://doi.org/10.1016/j.physletb.2007.09.077. arXiv:0707.1378

68. M. Cacciari, G.P. Salam, G. Soyez, The catchment area of jets. JHEP **04**, 005 (2008). https://doi.org/10.1088/1126-6708/2008/04/005. arXiv:0802.1188

69. N.D. Christensen, P. de Aquino, C. Degrande, C. Duhr, B. Fuks, M. Herquet et al., A comprehensive approach to new physics simulations. Eur. Phys. J. C **71**, 1541 (2011). https://doi.org/10.1140/epjc/s10052-011-1541-5. arXiv:0906.2474

70. A. Alloul, N.D. Christensen, C. Degrande, C. Duhr, B. Fuks, FeynRules 2.0: a complete toolbox for tree-level phenomenology. Comput. Phys. Commun. **185**, 2250–2300 (2014). https://doi.org/10.1016/j.cpc.2014.04.012. arXiv:1310.1921

71. C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, T. Reiter, UFO: the universal FeynRules output. Comput. Phys. Commun. **183**, 1201–1214 (2012). https://doi.org/10.1016/j.cpc.2012.01.022. arXiv:1108.2040

72. J.A. Aguilar-Saavedra, B. Fuks, M.L. Mangano, Pinning down top dipole moments with ultra-boosted tops. Phys. Rev. D **91**, 094021 (2015). https://doi.org/10.1103/PhysRevD.91.094021. arXiv:1412.6654