Studies of Boosted Topologies and Jet Substructure at the LHC

Johannes Haller[1], *Roman Kogler*[1], *Frank J. Tackmann*[2]

[1]Institut für Experimentalphysik, Universität Hamburg, Germany
[2]DESY, Hamburg, Germany

DOI: http://dx.doi.org/10.3204/PUBDB-2018-00782/B2a

The high centre-of mass energy available in run-2 of CERN's Large-Hadron-Collider (LHC) results in an abundant production of heavy particles with high transverse momenta (p_T). The boosted decay products of these high-p_T particles are collimated in the laboratory system and, in case of hadronic decays, the resulting jets are merged into a single larger jet which features certain characteristics in its substructure. During the last decade, several novel techniques have been developed to study and identify these boosted hadronic decays of heavy particles at the LHC. In this article, we summarize the results of developments on boosted topologies and jet substructure as obtained in the context of the SFB 676 in Hamburg and we discuss a few example results of their application.

1 Introduction

The large LHC datasets collected at $\sqrt{s} = 13$ TeV over the past years, open the possibility to study the yet unexplored region of very high invariant masses of new resonances. An important feature of many BSM scenarios is the abundant production of heavy SM particles (t, W, Z and H) as decay products of these new high-mass resonances. As a result, an increased production rate of these particles compared to the SM prediction can provide direct evidence for new physics. Furthermore, the study of the production of Higgs bosons at large transverse momenta, leading to boosted Higgs decays, is an important ingredient in the determination of properties and couplings of the Higgs boson

The high transverse momentum of the decaying t, W, Z, or H lead to boosted topologies in the final state with non-isolated leptons and merged jets. In the extreme case, all decay products of the decay chain are merged into a single large jet. In classical analyses, these final states cannot be identified and boosted heavy-particle decays cannot be distinguished from the large background of QCD multijet production. For an increased sensitivity, dedicated techniques for the reconstruction of such boosted topologies are required. This implies the development of new methods for the reconstruction and identification of boosted heavy objects, as well as their theoretical description. The key idea, which has long been recognized, is the study of the substructure of the large jets. Jet substructure techniques have become increasingly important both in the standard model as well as in searches for new physics. Often the techniques exploit the properties of a fixed number of subjets. Theoretical predictions at increasingly high precision are needed to match the increasing precision of the data, but are challenging due to the appearance of multiple widely separate physical scales, requiring the development of new factorization and resummation techniques.

We summarize the results in the area of boosted topologies and jet substructure obtained in the context of the SFB 676 in Hamburg. We present the concepts of the first measurement of an important substructure observable, the distribution of the jet mass in highly boosted $t\bar{t}$ events as performed by the CMS collaboration, followed by an overview of the key features of the HOTVR tagging algorithm. We highlight the theoretical developments of exploiting the N-jettiness observable as a factorization and resummation friendly exclusive cone jet algorithm, and we summarize the developments of novel effective field theories that enable the factorization and resummation of large logarithms appearing in boosted regimes and in the jet-mass spectrum. Finally, example BSM search results of the CMS collaboration are discussed exploiting the techniques of tagging and substructure observables.

2 Measurement of the jet mass in boosted $t\bar{t}$ events

One of the most important jet substructure observables is the jet mass m_{jet}, the invariant mass calculated from the jet constituents[1]. The jet mass of a large jet with a fully merged, highly boosted decay of t, W, Z or H is subject to various contributions, like multiple proton-proton scatterings in one bunch crossing (pileup), additional partonic radiation, multiple parton interactions and hadronisation effects. A detailed understanding and the measurement of the m_{jet} distribution is therefore crucial to test important components of the simulation, needed for LHC analyses of highly boosted topologies in the SM and for searches of new physics. In the case of fully hadronic top quark decays, the distribution of m_{jet} also provides sensitivity to the top quark mass m_t.

Calculations from first principles of the m_{jet} distribution for fully-merged top quark decays are available in soft collinear effective theory in e^+e^- collisions [2,3] and most recently for the LHC environment [4–6]. As a result, the measurement of the m_{jet} distribution can be used for an independent m_t determination in the boosted regime with the aim of reaching a reliable correspondence between the top quark mass in any well-defined renormalisation scheme and the top quark mass parameter in general-purpose event generators.

While measurements of m_{jet} at particle level had been done by the ATLAS and CMS collaborations for light-quark and gluon jets [7,8], the measurement of the differential $t\bar{t}$ production cross section as a function of the leading-jet mass for boosted top quark decays has been done for the first time in [1] using the $\sqrt{s} = 8\,\mathrm{TeV}$ data-set recorded by the CMS collaboration. The measurement is performed on $t\bar{t}$ events in which the leading jet includes the three subjets of one hadronic top quark decay, while the other top quark decays in the $l\nu_l b$ mode. The Cambridge–Aachen (CA) jet-clustering algorithm [9,10] with a distance parameter of $R = 1.2$ and $p_T > 400\,\mathrm{GeV}$ is used for the definition of the highly boosted top quark jets used in the measurement.

In Fig. 1 the m_{jet} distribution unfolded to the particle level and normalised to the fiducial-region total cross section as obtained from the CMS data is shown and compared to the equivalent distributions from Monte-Carlo (MC) simulations. The data distribution agrees with predictions from simulations indicating a high quality of the modelling of the jet mass in the simulation of highly boosted top quark decays. The overall sensitivity of a possible mass measurement derived from this measurement has been determined from a comparison of the data distribution with distributions obtained from simulations with different values of m_t and a

[1] The text presented in this section is based on the text of [1].

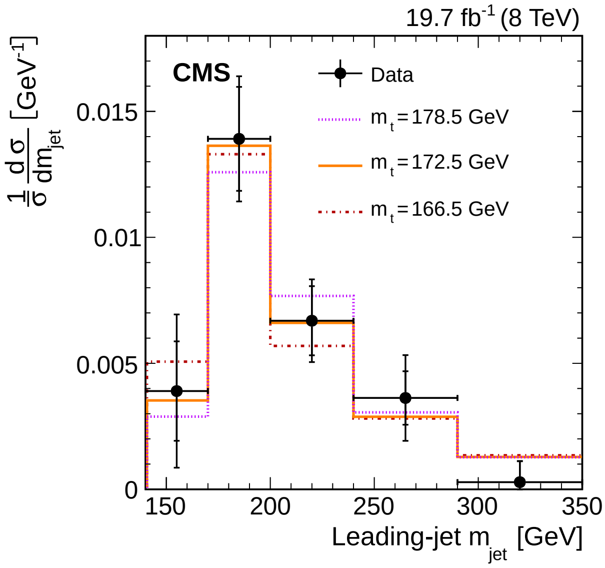

Figure 1: Normalised particle-level $t\bar{t}$ differential cross section as a the function of the leading-jet mass as measured by the CMS collaboration using pp collisions at $\sqrt{s} = 8\,\text{TeV}$. The CMS measurement (data points) is compared to MC predictions for three different values of m_t (solid, dotted and dashed histograms). Figure taken from Ref. [1].

sensitivity of $\pm 9\,\text{GeV}$ has been obtained. The dominating uncertainty results from the limited statistics ($\pm 6.0\,\text{GeV}$), while the experimental systematic uncertainty is $\pm 2.8\,\text{GeV}$. The modelling uncertainty contributes $\pm 4.6\,\text{GeV}$, while the theoretical uncertainty due to missing higher orders in the simulation amounts to $\pm 4.0\,\text{GeV}$.

While the current precision of a m_t determination using this method is not comparable with the precision obtained by other mass measurements, the analysis represents a proof-of-principle of the method. With more data collected at higher centre-of-mass energies the dominating statistical uncertainty of this measurement will drastically improve. In addition, the experimental systematic uncertainties will be significantly reduced through improved techniques in the reconstruction of boosted hadronic top quark decays [11, 12]. Furthermore, the improved statistical precision will allow for stronger constraints on the simulation of boosted top decays, leading to improvements of the modelling uncertainties. Lastly, new generations of MC event generators include full next-to-leading order calculations, which reduce the theoretical uncertainties. Overall, competitive top quark mass measurements using this technique can be envisaged for future data analyses by the LHC collaborations.

3 Identification of heavy hadronically decaying particles at the LHC with the HOTVR tagger

Since the start of the LHC a variety of innovative approaches based on the substructure of jets have been developed to identify decays of heavy hadronically decaying particles from the overwhelming background of QCD multijet production (see [13] for a list of references)[2]. Most of the developed algorithms are designed for a specific kinematic region (sometimes following an involved algorithmic procedure): either they are optimized for the region of low transverse momentum p_{T}, where the decay products can be resolved, or they provide an optimized performance in the boosted regime of high p_{T}.

During the course of the SFB 676 we developed a new algorithm [14] for the identification of boosted, hadronically decaying, heavy particles which is based on the procedure of jet clustering with variable distance parameter R [15] to adapt the jet size to jet transverse momentum p_{T}. It is the first application of the variable-R jet clustering algorithm for tagging boosted heavy SM decays. The advantage of a shrinking cone size with increasing p_{T} is that contributions of perturbative and non-perturbative effects on jet substructure observables increase with p_{T}, and are compensated by a decreasing jet size. Additionally, the algorithm identifies subjets using a mass jump condition [16]. The resulting algorithm, Heavy Object Tagger with Variable R (HOTVR) [14], has only little algorithmic complexity. It performs the clustering of jets, the identification of subjets and the rejection of soft clusters in a single sequence. While the HOTVR algorithm is in principle applicable for the tagging of any heavy hadronically decaying particle (W, Z, H, t or possible BSM resonances) the focus of past studies [14] has been put on performance studies of the identification of top quark decays for which selection cuts on measureable substructure observables have been studied and optimized.

Certain exemplary features of the performance of the algorithm are illustrated in Fig. 2 where the results of the application of the HOTVR algorithm are shown for two simulated $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ at low p_{T} (left) and at high p_{T} (right). The HOTVR algorithm indeed adapts the jet size to the jet p_{T} and it is able to reconstruct correctly the sub-jets (orange/blue) within the larger jets, thereby identifying the initial top quark decay products (red circles) and rejecting soft clusters originating from additional QCD radiation (grey areas) with the mass jump criterion.

Compared to other top tagging algorithms used within the CMS collaboration [17], the HOTVR algorithm was found [14] to demonstrate a remarkably stable performance in a wide range of top quark p_{T}, such that it can be used to cover the regions from nearly resolved decays up to the highly boosted regime with very high efficiency. As result, the HOTVR tagger could be used at the LHC in searches for hypothetical BSM particles decaying to top quarks in the low, the intermediate and the high mass region simultaneously.

While the performance studies of the HOTVR have focused on decays of the top quark so far, the identification of W, Z and H decays or hypothetical BSM resonances remains a potential subject for future studies in this area. Because of its stable performance over a wide range of different kinematic regions and its algorithmic simplicity, the HOTVR tagger will become a helpful ingredient for future boosted analyses at the LHC.

4 N-jettiness as a jet algorithm

As a global event shape, N-jettiness measures the degree to which the hadrons in the final state are aligned along N jet axes or the beam direction [18]. It was originally introduced to veto additional jets in an event, providing a way to define and resum exclusive N-jet cross sections [18–20]. N-jettiness was later adapted to the jet shape N-subjettiness [21, 22], which is an efficient measure to identify N-prong boosted hadronic objects such as top quarks, W/Z

[2] The text presented in this section closely follows the text of [14].

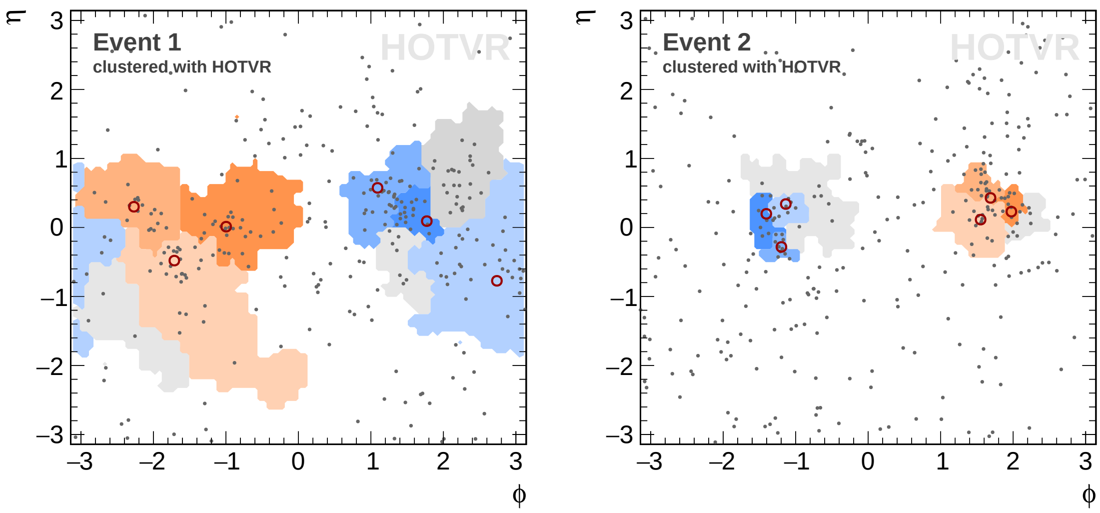

Figure 2: Reconstruction and (sub-)jet identification of simulated $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ at low p_T (left) and at high p_T (right) respectively using the HOTVR algorithm. The locations in the plane of the pseudo-rapidity η versus the azimuthal angle ϕ of the two leading jets in the events are shown as coloured areas (orange/blue). The locations of the stable particles (quarks from the top quark decay) are illustrated by grey dots (red circles). The subjets are shaded from light to dark, corresponding to increasing p_T. The grey areas correspond to regions rejected by the mass jump criterion. Figures taken from Ref. [14].

bosons, and Higgs bosons within a larger jet. By minimizing N-(sub)jettiness, one can directly identify N (sub)jet directions. By now, N-subjettiness has become an important jet substructure observable.

While hadronic jets are crucial to connect the observed hadronic final state to the short-distance hard interaction, the definition of a hadronic jet is ambiguous, as there is no unique way to map colour-singlet hadrons to colour-carrying partons. Moreover, different physics applications can benefit from different jet definitions. For these reasons, a wide variety of jet algorithms have been proposed, though currently, most LHC measurements involve jets clustered with the anti-k_T algorithm [23].

The possibility of using N-jettiness as a jet algorithm was already pointed out in [18]. In [12], a new jet algorithm called "XCone" is developed. It is based on minimizing the event shape N-jettiness [18] and uses developments from the jet shape N-subjettiness [21, 22]. The key feature is that N-jettiness defines an *exclusive cone* jet algorithm, i.e, it always returns a predetermined fixed number of jets, relevant for physics applications where the number of jets is known in advance. Like anti-k_T jets, XCone jets are nearly conical for well-separated jets. Typically, when using other jet algorithms, the boosted regime of overlapping jets requires separate analysis strategies using large jets with substructure. In contrast, with XCone the jets remain resolved even when jets are overlapping in the boosted regime. As a result, a key feature of XCone is that it smoothly transitions between the resolved regime where the N signal jets of interest are well separated and the boosted regime where they overlap. The returned value of N-jettiness also provides a quality criterion of how N-jet-like the event looks. In [11], examples

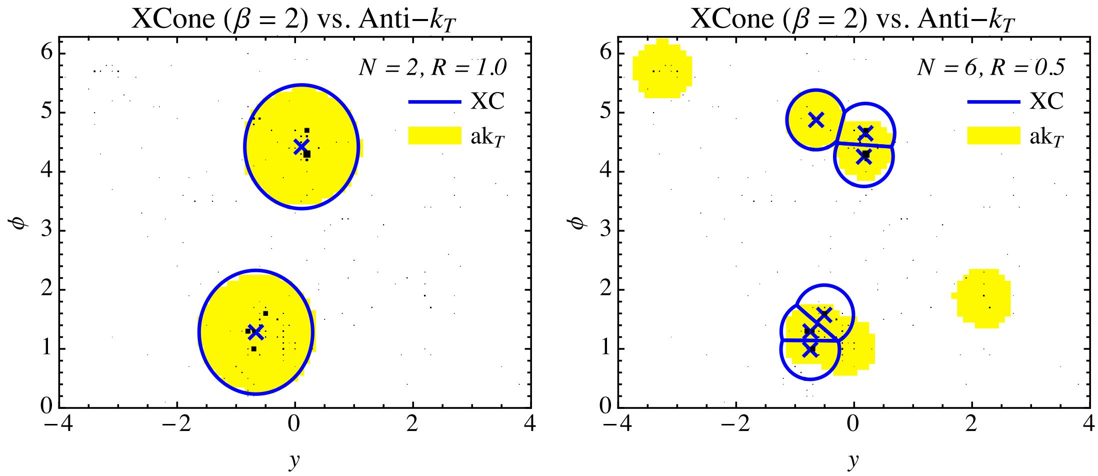

Figure 3: Comparison between the XCone (with its default measure) and anti-k_T for a boosted hadronic $t\bar{t}$ event. Left: Asking for two large-radius jets, the event is characterized by two widely-separated jets. For such jets, XCone yields nearly identical jet regions to anti-k_T. Right: Asking for six small-radius jets, the boosted hadronic top decay topology becomes clearly visible with XCone. Unlike anti-k_T which merges jet regions closer in angle than $\approx R$, XCone allows such jet regions to remain split, making it much more robust against picking up unrelated ISR jets. Figures taken from Ref. [12].

of quasi-boosted kinematics that capitalize on this feature are discussed for dijet resonances, Higgs decays to bottom quarks, and all-hadronic top-quark pairs.

There is considerable flexibility in precisely how one defines N-jettiness, and several different N-jettiness measures yielding different jet regions have been considered before [18, 20–22, 24]. The XCone default is a conical geometric measure that incorporates the insights from the different previous use cases. This measure is based on the dot product between particles and lightlike axes but incorporates an angular exponent β, as well as a beam exponent γ for additional flexibility. Most importantly, it is linear in the particle momenta, which simplifies theoretical calculations, as discussed below. At the same time, and crucially for the purposes of jet finding at the LHC, this measure yields conical jets over a wide rapidity range. Well separated jets are bounded by circles of radius R in the rapidity-azimuth plane, and are nearly identical to anti-k_T jets. This is illustrated in the left panel of Fig. 3. On the other hand, overlapping jet regions automatically form nearest-neighbour clover-like jets, as show in the right panel of Fig. 3. In contrast to anti-k_T, XCone does not merge the overlapping jets even in the boosted regime, making the 3-prong structure of the boosted top decay clearly visible.

The presence of jets makes perturbative QCD calculations challenging and lead to a complicated singularity structure. Furthermore, imposing a fixed number of jets through some kind of direct or indirect veto on additional jets restricts the phase space for additional collinear and soft emissions. This generates logarithms that often dominate the perturbative series and need to be resummed to obtain predictions with the best possible precision. Soft Collinear Effective Theory (SCET) [25–28] provides a framework to systematically carry out the resummation of logarithms to higher orders by factorizing the cross section into hard, collinear, and soft

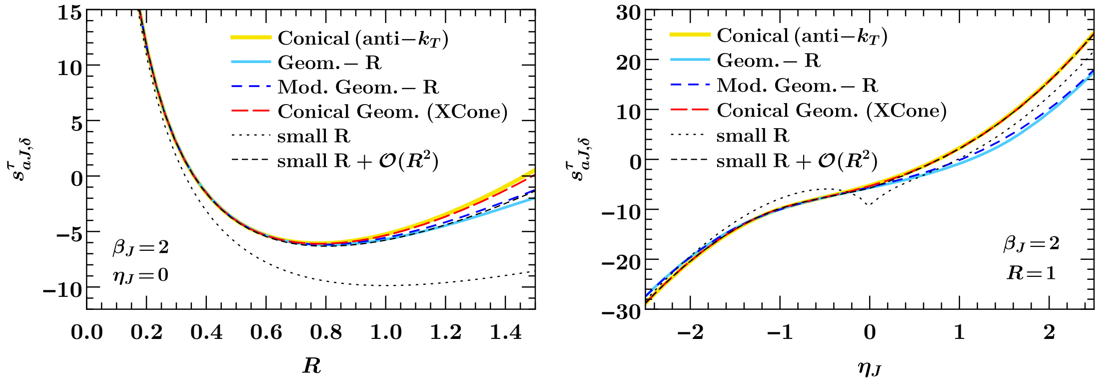

Figure 4: Soft coefficient describing initial-final state soft interference. Left: Jet radius dependence for different jet algorithm distance measures. The dotted and dashed curves show the small-R limit and including the first $O(R^2)$ corrections. Right: Dependence on the jet rapidity for fixed $R = 1$. Figures taken from Ref. [30].

functions, and then exploiting their renormalization group evolution. Schematically, the cross section for $pp \to N$ jets factorizes for many observables in the singular limit as [18, 19]

$$\sigma_N = H_N \times \left[B_a B_b \prod_{i=1}^{N} J_i \right] \otimes S_N , \tag{1}$$

where the hard function H_N contains the virtual corrections to the partonic hard scattering process, the beam functions $B_{a,b}$ contain parton distribution functions and describe collinear initial-state radiation. The jet functions J_i describe final-state radiation collinear to the direction of the hard partons, and the soft function S_N describes wide-angle soft radiation. The resummation of large logarithms is achieved by evaluating each component at its natural scale and then renormalization-group evolving all components to a common scale. The jet and beam functions typically do not depend on the precise definition of the jet regions and algorithm and are known for a variety of jet and beam measurements. Hard functions are also known for many processes at one loop or beyond (see e.g. Ref. [29] and references therein).

In [12, 30], the factorization theorems relevant for various N-jettiness measures, including the XCone default, and jet measurements are determined. The resummation at NNLL requires the soft function at one loop. Compared to the beam and jet functions, the perturbative calculation of the soft function generally requires a more sophisticated setup, since it depends not only on the measurements made in the jet and beam regions, but also on the angles between all jet and beam directions and the precise definition of the jet boundaries, i.e., the considered jet algorithm. In [30] a general method is developed to compute the N-jet soft functions for a wide class of jet algorithms and jet measurements, which so far have been known only for specific cases [20, 31]. Different types of jet vetoes are discussed, including beam thrust, beam C parameter, and a jet-p_{T} veto. For the jet algorithm of the signal jets, different partitionings are considered, including anti-k_T and XCone. It is found that the one-loop soft function can be written in terms of universal analytic contributions and a set of numerical integrals, which explicitly depend on the partitioning and observable. Fully analytical results can be obtained in the limit of small jet radius R. Interestingly, it is found that the small-R expansion works remarkably well for the soft function even for moderate values of R, if one includes corrections up to $O(R^2)$. This is illustrated in Fig 4, which shows as an example the contribution to the soft function from soft gluon interference between the incoming and outgoing hard partons.

5 Factorization in boosted regimes and at small jet radius

In multijet events one generically encounters additional hierarchies in the hard kinematics of the jets, namely among the jet energies and/or among the angles between jets. At the LHC, an important class of examples are precisely jet substructure methods to reconstruct boosted heavy objects, which essentially rely on identifying soft or collinear (sub)jets. Another example is cascade decays of heavy new (coloured) particles leading to experimental signatures with jets of widely different p_{T}. There are also cases where additional jets produced by QCD are used to tag or categorize the signal events, a prominent example being the current Higgs measurements. Whenever such kinematic hierarchies arise among QCD-induced jets, in particular in the corresponding background processes, the enhancement of soft and collinear emissions in QCD leads to additional logarithms of the jet kinematics in the cross section. So far, a complete and general factorization framework for multijet processes that allows for a systematic resummation of such kinematic logarithms for generic jet hierarchies has been missing. Current predictions therefore rely on Monte Carlo parton showers and are thus mostly limited to leading-logarithmic (LL) accuracy.

In [32], we developed a general factorization framework called SCET$_+$, which is an extension of standard SCET and allows for a systematic higher-order resummation of such kinematic logarithms for generic jet hierarchies. Compared to the usual soft and collinear emissions present in SCET, in SCET$_+$ the effects of additional intermediate emissions are considered that have both soft and collinear characteristics and describe the production and interaction of hierarchical soft and collinear jets. The resulting factorized cross sections amount to further factorizing the hard and soft functions appearing in the non-hierarchical case in Eq. (1), with the additional factorization ingredients given in terms of collinear splitting amplitudes and soft gluon currents, which fully capture spin and colour correlations.

Special cases have been considered before, in particular SCET$_+$ first appeared in Ref. [33], where its purely collinear regime was constructed to describe the situation of two energetic jets collinear to each other. The purely soft regime of SCET$_+$ was first considered in Ref. [34]. There it was shown that this regime is essential for the resummation of non-global logarithms by explicitly resolving additional soft subjets (see also Refs. [35, 36]).

In [32], the general application to N-jet processes at hadron colliders is derived in detail, considering all relevant representative classes of hierarchies, from which the general case can be built, and it is discussed how to systematically combine the different hierarchical (boosted) and non-hierarchical regimes to obtain a complete description of the kinematic jet phase space. This includes in particular multiple hierarchies that are either strongly ordered in angle or energy or not. Our results pave the way to resum kinematic logarithms appearing in differential jet measurements to NNLL or beyond and systematically improve upon the LL description of kinematic logarithms in parton showers.

As mentioned before, the same type of jet hierarchies also appear in jet substructure measurements, and SCET$_+$ has become the cornerstone of analytic resummation of jet substructure observables. For example, in Ref. [35], SCET$_+$ was used to factorize and resum a two-prong jet

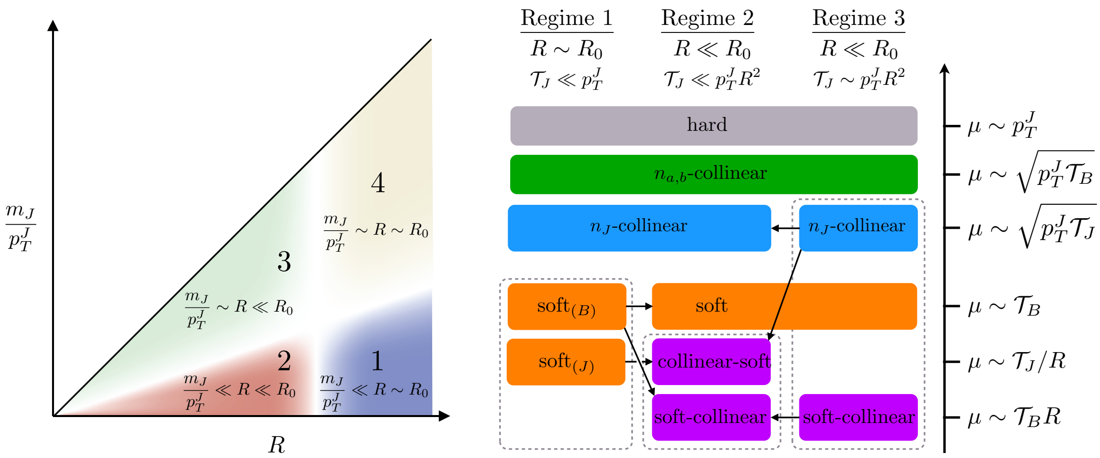

Figure 5: Illustration of the various regimes for jet mass measurements as a function of the jet radius R and p_T^J/m_J (left panel) and the resulting physical scales and radiation types that must be considered in the factorization (right panel). Figures taken from Ref. [42].

substructure variable and in Ref. [37] it was used to provide the first resummed hadron-level predictions for a groomed 2-prong jet substructure variable.

A basic and important benchmark observable for studying the radiation inside a jet is the invariant mass m_J of a jet. The jet mass spectrum provides key information about the influence of Sudakov double logarithms and soft radiation in a hadronic environment and probes the dependence on the jet algorithm and jet size R, colour flow, initial and final state partonic channels, hadronization, and underlying event. For this reason, it is a key benchmark observable for jet substructure studies. For example, the utility of the first moment of the jet mass spectrum as a mechanism to disentangle different sources of soft radiation underlying the hard interaction was discussed in Ref. [38]. The best sensitivity to these effects comes from studying jets in their primal state, without using jet-grooming techniques to change the nature of the jet constituents. While useful for tagging studies, jet grooming fundamentally changes the nature of the jet mass observable and reduces its utility as a probe of these physical effects.

To predict the jet mass spectrum at a hadron collider, it is crucial to account for the resummation of logarithms between the transverse momentum p_T^J of the jet and its invariant mass m_J. The jet-mass spectrum for N-jettiness jets (defined via the standard geometric N-jettiness measure) was first calculated at NNLL in [24]. Other analytic ungroomed jet-mass calculations have been carried out in Refs. [38–41]. In [42], the analytic description of exclusive jet mass spectra at the LHC is extended to realistic jet algorithms, including both anti-k_T clustering and XCone. A particular focus is on the limit of small jet radius, which turns out to be relevant already for moderately small $R \lesssim 1$ as typically used in experiments. In particular, for small R, the exclusive N-jet cross section contains Sudakov double logarithms of R, in conjunction with logarithms of the jet mass and jet veto. It turns out that the factorization in the small-R limit provides a different application of SCET$_+$, which was already found for the case of jet rates (without a measurement on the jet) in Refs. [43, 44].

As illustrated in Fig. 5, one must in principle distinguish four different regimes with different hierarchies for R and the scales m_J and $p_T^J R$:

- regime 1: large-R jets for small m_J: $m_J \ll p_T^J R \sim p_T^J$
- regime 2: small-R jets for small m_J: $m_J \ll p_T^J R \ll p_T^J$
- regime 3: small-R jets for large m_J: $m_J \sim p_T^J R \ll p_T^J$
- regime 4: large-R jets for large m_J: $m_J \sim p_T^J R \sim p_T^J$.

All of these require distinct effective field theory setups to resum the corresponding logarithms. Specifically, in regimes 1 and 2 these are logarithms of m_J/p_T^J, and in regimes 2 and 3 logarithms of R. The factorization for each of the different regimes is discussed, as well as the relations between the different regimes and how to combine them. A particular focus is on regime 2, which has the most phenomenologically interest. Here, by applying SCET$_+$, all relevant scales associated with m_J, R, the applied jet veto, and the p_T^J of the jet are fully factorized, enabling the systematic resummation of jet-radius logarithms in addition to the jet-mass logarithms beyond leading-logarithmic order.

6 Application of boosted techniques in the search for BSM physics at the LHC

From the start of data taking at the LHC, boosted techniques have become a very important tool in searches for new physics. The reason is the much larger accessible resonance mass range compared to previous colliders, which results in more collimated final states. A large number of searches have benefitted from the application of jet substructure algorithms, either in terms of improved sensitivity or by making the search feasible in channels which are not accessible with standard reconstruction methods. In this section, we discuss a few results obtained by the CMS collaboration by using these techniques in the search for BSM physics.

An important example is the search for resonant $t\bar{t}$ production, which is predicted to be mediated by BSM particles with masses of multiple TeVs in a number of new physics models. Analyses that have been carried out during the course of the SFB 676 use the full $\sqrt{s} = 8\,\text{TeV}$ dataset [45, 46] and data with $\sqrt{s} = 13\,\text{TeV}$ recorded in the years 2015 [47] and 2016 [48] by the CMS collaboration. The analyses consider leptonic and fully-hadronic final states. The requirement of a jet from a boosted top quark decay in the lepton+jets channel results in a considerable improvement in sensitivity, compared to traditional reconstruction methods. The analyses of all-hadronic channels, optimized for low and high resonance masses, were only made possible through jet substructure methods, which achieved an impressive reduction of the SM multijet background by more than four orders of magnitude. An crucial aspect of these searches is the determination of the SM multijet background from data since the simulation of multi-parton final states with light quarks and gluons produced through the strong force is not precise enough for the demands of these searches. Jet substructure methods offer a reliable possibility for the determination of these backgrounds through control regions. These can be obtained by inverting criteria on the jet substructure observables used for identifying boosted top quark decays. The jets measured in these control regions have very similar kinematic distributions as the ones in the signal regions, resulting in reliable extrapolations of the SM backgrounds with small uncertainties. The analyses still rely on an accurate modelling of jet substructure observables for the determination of irreducible backgrounds from SM $t\bar{t}$ production and the signal efficiencies.

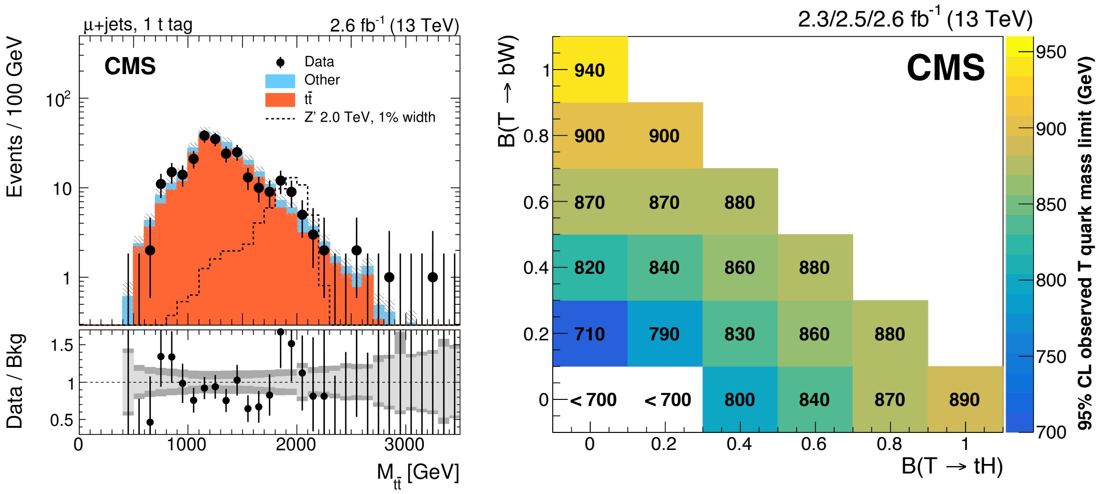

Figure 6: Example results of CMS searches using boosted top quark reconstruction techniques in pp collisions at $\sqrt{s} = 13\,\text{TeV}$: (left) Distribution of the $t\bar{t}$ invariant mass in μ+jets events including a top-tagged jet in the search of $t\bar{t}$ resonances. Shown are the data, the expected SM background contribution as well as the BSM signal normalized to a signal cross section of 1 pb. (right) Observed 95 % CL exclusion limit on the mass of vector-like T quark for a variety of $T \to tH$ and $T \to bW$ branching fraction combinations. Figures taken from Refs. [47] and [49], respectively.

The measured distribution of the mass of the $t\bar{t}$ system in the muon+jets channel with an identified boosted top quark decay is shown in Fig. 6 (left). The dominant background is irreducible SM $t\bar{t}$ production, and a possible BSM signal would be visible as a peak on top of the falling background distribution. The power of jet substructure methods in searches for new physics is demonstrated when comparing the sensitivity of the individual channels. At high resonance masses, analyses in the all-hadronic channels achieve comparable sensitivity to the ones in the lepton+jets channels. When combined, these searches have placed the most stringent constraints on resonant $t\bar{t}$ production in pp collisions at the time of publication.

Another example class of BSM scenarios that can be probed with jet substructure methods are models with vector-like quarks (VLQs), which extend the three generations of SM quarks by a fourth generation with a different chiral structure. These VLQs are expected to be much heavier than the third generation quarks with masses not generated by the BEH mechanism and thus avoiding constraints from H coupling measurements. Possible decay modes of the VLQs are bW, tZ and tH for heavy partners of the top quark (T) and tW, bZ and bH for partners of the bottom quark (B). Pair production of heavy VLQs at the LHC would thus result in complex final states with decays of highly boosted t, W, Z and H.

A search by the CMS collaboration targeting the final states bW and tH has been carried out using data recorded in 2015 with $\sqrt{s} = 13$ TeV [49]. The decay of one VLQ is reconstructed through the leptonic decay of a W boson, thus reducing the SM multijet background. The presence of a second VLQ is surmised through jet substructure methods, requiring a jet with substructure compatible to the fully-merged hadronic decay of a W or H boson. The results of the analysis in terms of VLQ mass exclusion limits are shown in Fig. 6 (right). The sensitivity of this search could be improved considerably by advanced techniques to identify the highly-boosted W and H decays. This result has been a precursor for a number of VLQ searches using data with $\sqrt{s} = 13$ TeV.

7 Conclusion

In this contribution, we have summarized the results that have been obtained in the context of the SFB 676 in the area of studies of boosted topologies and jet substructure at the LHC. Example results are concepts for the development of the first measurements of the jet mass distribution in boosted $t\bar{t}$ events by the CMS collaboration, the development of the HOTVR tagger, theoretical developments of the exploitation of the N-jettiness observable as a jet algorithm and developments of novel effective field theories that enable the factorization and resummation of large logarithms appearing in boosted regimes and in the jet-mass spectrum. In addition, we discussed a few example results of searches for new physics by the CMS collaboration using boosted reconstruction techniques.

Acknowledgments

J.H. and R.K. would like to thank their collaborators in the CMS collaboration. The work presented here has been supported in parts by the DFG through the SFB 676 "Particles, Strings and the Early Universe" located in Hamburg.

References

[1] CMS Collaboration, A. M. Sirunyan et al., *Measurement of the jet mass in highly boosted $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C77** (2017) 467, [1703.06330].

[2] S. Fleming, A. H. Hoang, S. Mantry and I. W. Stewart, *Jets from massive unstable particles: Top-mass determination*, *Phys. Rev.* **D77** (2008) 074010, [hep-ph/0703207].

[3] S. Fleming, A. H. Hoang, S. Mantry and I. W. Stewart, *Top Jets in the Peak Region: Factorization Analysis with NLL Resummation*, *Phys. Rev.* **D77** (2008) 114003, [0711.2079].

[4] A. H. Hoang, A. Pathak, P. Pietrulewicz and I. W. Stewart, *Hard Matching for Boosted Tops at Two Loops*, *JHEP* **12** (2015) 059, [1508.04137].

[5] M. Butenschoen, B. Dehnadi, A. H. Hoang, V. Mateu, M. Preisser and I. W. Stewart, *Top Quark Mass Calibration for Monte Carlo Event Generators*, *Phys. Rev. Lett.* **117** (2016) 232001, [1608.01318].

[6] A. H. Hoang, S. Mantry, A. Pathak and I. W. Stewart, *Extracting a Short Distance Top Mass with Light Grooming*, 1708.02586.

[7] ATLAS Collaboration, G. Aad et al., *Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*, *JHEP* **05** (2012) 128, [1203.4606].

[8] CMS Collaboration, S. Chatrchyan et al., *Studies of jet mass in dijet and W/Z + jet events*, *JHEP* **05** (2013) 090, [1303.4811].

[9] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[10] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279, 1998, hep-ph/9907280.

[11] J. Thaler and T. F. Wilkason, *Resolving Boosted Jets with XCone*, *JHEP* **12** (2015) 051, [1508.01518].

[12] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion and T. F. Wilkason, *XCone: N-jettiness as an Exclusive Cone Jet Algorithm*, *JHEP* **11** (2015) 072, [1508.01516].

[13] L. Asquith et al., *Jet Substructure at the Large Hadron Collider : Experimental Review*, 1803.06991.

[14] T. Lapsien, R. Kogler and J. Haller, *A new tagger for hadronically decaying heavy particles at the LHC*, *Eur. Phys. J.* **C76** (2016) 600, [1606.04961].

[15] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) 059, [0903.0392].

[16] M. Stoll, *Vetoed jet clustering: The mass-jump algorithm*, *JHEP* **04** (2015) 111, [1410.4637].

[17] CMS Collaboration, *Top Tagging with New Approaches*, *CMS-PAS-JME-15-002* (2016) . http://cds.cern.ch/record/2126325.

[18] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002, [1004.2489].

[19] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Factorization at the LHC: From PDFs to Initial State Jets*, *Phys. Rev.* **D81** (2010) 094035, [0910.0467].

[20] T. T. Jouttenus, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *The Soft Function for Exclusive N-Jet Production at Hadron Colliders*, *Phys. Rev.* **D83** (2011) 114030, [1102.4344].

[21] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [1011.2268].

[22] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093, [1108.2701].

[23] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[24] T. T. Jouttenus, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Jet mass spectra in Higgs boson plus one jet at next-to-next-to-leading logarithmic order*, *Phys. Rev.* **D88** (2013) 054031, [1302.0846].

[25] C. W. Bauer, S. Fleming and M. E. Luke, *Summing Sudakov logarithms in B —> X(s gamma) in effective field theory*, *Phys. Rev.* **D63** (2000) 014006, [hep-ph/0005275].

[26] C. W. Bauer, S. Fleming, D. Pirjol and I. W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, *Phys. Rev.* **D63** (2001) 114020, [hep-ph/0011336].

[27] C. W. Bauer and I. W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett.* **B516** (2001) 134–142, [hep-ph/0107001].

[28] C. W. Bauer, D. Pirjol and I. W. Stewart, *Soft collinear factorization in effective field theory*, *Phys. Rev.* **D65** (2002) 054022, [hep-ph/0109045].

[29] I. Moult, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Employing Helicity Amplitudes for Resummation*, *Phys. Rev.* **D93** (2016) 094003, [1508.02397].

[30] D. Bertolini, D. Kolodrubetz, D. Neill, P. Pietrulewicz, I. W. Stewart, F. J. Tackmann et al., *Soft Functions for Generic Jet Algorithms and Observables at Hadron Colliders*, *JHEP* **07** (2017) 099, [1704.08262].

[31] T. Kasemets, W. J. Waalewijn and L. Zeune, *Calculating Soft Radiation at One Loop*, *JHEP* **03** (2016) 153, [1512.00857].

[32] P. Pietrulewicz, F. J. Tackmann and W. J. Waalewijn, *Factorization and Resummation for Generic Hierarchies between Jets*, *JHEP* **08** (2016) 002, [1601.05088].

[33] C. W. Bauer, F. J. Tackmann, J. R. Walsh and S. Zuberi, *Factorization and Resummation for Dijet Invariant Mass Spectra*, *Phys. Rev.* **D85** (2012) 074006, [1106.6047].

[34] A. J. Larkoski, I. Moult and D. Neill, *Non-Global Logarithms, Factorization, and the Soft Substructure of Jets*, *JHEP* **09** (2015) 143, [1501.04596].

[35] A. J. Larkoski, I. Moult and D. Neill, *Analytic Boosted Boson Discrimination*, *JHEP* **05** (2016) 117, [1507.03018].

[36] D. Neill, *The Edge of Jets and Subleading Non-Global Logs*, 1508.07568.

[37] A. J. Larkoski, I. Moult and D. Neill, *Factorization and Resummation for Groomed Multi-Prong Jet Shapes*, *JHEP* **02** (2018) 144, [1710.00014].

[38] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Dissecting Soft Radiation with Factorization*, *Phys. Rev. Lett.* **114** (2015) 092001, [1405.6722].

[39] M. Dasgupta, K. Khelifa-Kerfa, S. Marzani and M. Spannowsky, *On jet mass distributions in Z+jet and dijet processes at the LHC*, *JHEP* **10** (2012) 126, [1207.1640].

[40] Y.-T. Chien, R. Kelley, M. D. Schwartz and H. X. Zhu, *Resummation of Jet Mass at Hadron Colliders*, *Phys. Rev.* **D87** (2013) 014010, [1208.0010].

[41] Z. L. Liu, C. S. Li, J. Wang and Y. Wang, *Resummation prediction on the jet mass spectrum in one-jet inclusive production at the LHC*, *JHEP* **04** (2015) 005, [1412.1337].

[42] D. W. Kolodrubetz, P. Pietrulewicz, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Factorization for Jet Radius Logarithms in Jet Mass Spectra at the LHC*, *JHEP* **12** (2016) 054, [1605.08038].

[43] Y.-T. Chien, A. Hornig and C. Lee, *Soft-collinear mode for jet cross sections in soft collinear effective theory*, *Phys. Rev.* **D93** (2016) 014033, [1509.04287].

[44] T. Becher, M. Neubert, L. Rothen and D. Y. Shao, *Effective Field Theory for Jet Processes*, *Phys. Rev. Lett.* **116** (2016) 192001, [1508.06645].

[45] CMS Collaboration, S. Chatrchyan et al., *Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}$=8 TeV*, *Phys. Rev. Lett.* **111** (2013) 211804, [1309.2030]. [Erratum: Phys. Rev. Lett.112,no.11,119903(2014)].

[46] CMS Collaboration, V. Khachatryan et al., *Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev.* **D93** (2016) 012001, [1506.03062].

[47] CMS Collaboration, A. M. Sirunyan et al., *Search for $t\bar{t}$ resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2017) 001, [1704.03366].

[48] CMS Collaboration, *Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *CMS-PAS-B2G-17-017* (2018) . http://cds.cern.ch/record/2621631.

[49] CMS Collaboration, A. M. Sirunyan et al., *Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **11** (2017) 085, [1706.03408].