学位論文

New methods for top quark identification and reconstruction at hadron colliders

（ハドロンコライダーにおけるトップクォークの同定と再構成
のための新しい手法）

平成２７年７月博士（理学）申請

東京大学大学院理学系研究科
物理学専攻
シュトル　マルティン

New methods for top quark identification and reconstruction at hadron colliders

Martin Stoll

Department of Physics, Faculty of Science
The University of Tokyo
Japan

A thesis submitted for the degree of Doctor of Philosophy in 2015.

July 10, 2015
Revised: August 27, 2015

For Yuka.

Preface

Main parts of this dissertation are based on the following three works by the author.

[A] M. Stoll, "Vetoed jet clustering: The mass-jump algorithm," *JHEP*, vol. 1504, p. 111, 2015. DOI: 10.1007/JHEP04(2015)111. arXiv: 1410.4637 [hep-ph].

[B] K. Hamaguchi, S. P. Liew, and M. Stoll, "Jumping into buckets, or How to decontaminate overlapping fat jets," *Phys. Rev.*, vol. D92, no. 1, p. 015 012, 2015. DOI: 10.1103/PhysRevD.92.015012. arXiv: 1505.02930 [hep-ph].

[C] M. Spannowsky and M. Stoll, "Tracking New Physics at the LHC and beyond," 2015. arXiv: 1505.01921 [hep-ph].

The following work by the author is referenced in this dissertation.

[D] M. Endo, K. Hamaguchi, K. Ishikawa, and M. Stoll, "Reconstruction of Vector-like Top Partner from Fully Hadronic Final States," *Phys.Rev.*, vol. D90, no. 5, p. 055 027, 2014. DOI: 10.1103/PhysRevD.90.055027. arXiv: 1405.2677 [hep-ph].

Code developed in the context of this dissertation is publically available from the following sources.

[E] *FastJet Contrib*, http://fastjet.hepforge.org/contrib/.

[F] *HPTTopTagger, HPTWTagger, and HPTZTagger*, https://www.ippp.dur.ac.uk/~mspannow/webippp/HPTTaggers.html.

Abstract

One of the main purposes of the Large Hadron Collider (LHC) is to investigate the mechanism behind electroweak symmetry breaking. The Standard Model top quark plays a key role in this task, as it has by far the largest coupling constant to the Higgs field. Due to the large centre-of-mass energy in run II of the LHC, processes involving top quarks with large kinetic energy will come into focus more than before. Furthermore, similar final states can be significantly enhanced if the top quark is the decay product of a yet undiscovered heavy resonance. In fact, many theories that address electroweak symmetry breaking introduce addition resonances at the TeV scale. These particles can subsequently decay into "boosted" t, H, and electroweak gauge bosons. In this thesis, we review state-of-the-art tagging algorithms for boosted top quarks and also develop new algorithms. Our guiding principle is the energy frontier accessible at the LHC.

We investigate pair production of vectorlike tops with TeV-scale mass. In our scenario, this process leads to a high-multiplicity final state with two boosted top quarks and two boosted Higgs bosons. Common tagging algorithms developed for well-isolated boosted particles struggle, and we apply new top and Higgs taggers based on a novel jet clustering algorithm. We find good reconstruction quality and significant improvement over benchmark analyses with standard present-day algorithms. The key ingredient is jet clustering with a terminating veto, which we dubbed "mass-jump" clustering. We also find merits under simpler, well-controlled circumstances, where we investigate similarities and differences between the novel mass-jump algorithm and commonly used tools.

If existent, very heavy resonances with masses of multiple TeV will also become accessible. Their Standard Model decay products can be so highly boosted that common reconstruction techniques seize to be applicable. For highly boosted top quarks, we refine a tagger based on charged particle tracks. We generalize the algorithm to highly boosted W and Z bosons as well. For all taggers we find good discrimination power and kinematic reconstruction quality. Implementing them in dedicated searches can expand the discovery reach of new heavy resonances significantly.

Both scenarios described above did not receive much attention in the literature so far, but are likely to become accessible in run II at the LHC. Our algorithms can play important roles in searches for new physics, pushing further the energy frontier.

Contents

List of Figures

List of Tables

— The unknown knowns. Things that you possibly may know that you don't know you know.

Donald Henry Rumsfeld

1 Introduction

Man's desire to discover and understand the mysteries the universe holds for us has driven philosophers throughout the ages. Yet what we consider modern science is based on reproducible experiments, or in other words observations of well-prepared systems that are intended to answer specific questions posed on Nature. Without technology to enhance our five senses, our conception of the world could not have taken the radical turns it experienced during the last centuries. Telescopes allow us to peak far into other galaxies and the very beginning of the universe, where on the other end of the spectrum energetic particle colliders empower us to study the very building blocks of matter beyond the subatomic scale. In order to probe smaller and smaller distances, the collision energy of these particle colliders has been increased steadily. As a result, now all fields of the Standard Model of Particle Physics (SM) are confirmed experimentally. The Standard Model merges electroweak theory of the electromagnetic and weak nuclear forces [1]–[3] with quantum chromodynamics of the strong nuclear force [4]–[7]. It can successfully explain the results of almost every experiment conducted so far to great precision and incorporates all observed matter and forces except gravity. Despite this triumph, a few unpleasantries do remain. Some of them are of the aesthetic kind like the large and "unnatural" [8] range of elementary masses and coupling constants, others are severe incompatibilities with experiment. The quest for the ultimate theory of nature is far from being over.

The Large Hadron Collider (LHC) operated at CERN in Geneva once again pushes the energy frontier to an unprecedented centre-of-mass energy of $\sqrt{s} = 13 \sim 14\,\mathrm{GeV}$. One of the main purposes of this machine is to solve the mystery of electroweak symmetry breaking [9]–[12], the very mechanism that lead to a Nobel Prize for François Englert and Peter W. Higgs in 2013. The particle bearing responsibility, the Higgs (sometimes labelled Brout-Englert-Higgs) boson, was found at the LHC [13], [14], but many

questions – whether it is indeed the particle of Standard Model, whether it is elementary or composed and held together by a new force, whether there are additional Higgs particles – all these questions remain unanswered.

Many theories beyond the Standard Model (abbreviated BSM) introduce new particles, many of which have escaped observation just because they are very heavy. We briefly review some models later in this thesis. When these new states are connected to the mechanism of electroweak symmetry breaking, they are expected to have a mass around the TeV scale if the theory wants to be labelled "natural" (see e.g. Refs. [15]–[17]). In the upcoming run, the LHC can and will penetrate this regime and put severe constraints on many such BSM models.

With the immense collision energy of the LHC, the window to more energetic final states is pushed wide open. In particular if the predicted heavy resonances at the TeV scale do exist, processes with large centre-of-mass energy will be significantly enhanced. The heavy resonances may decay into Standard Model particles around the electroweak scale and leave a footprint visible to the detector. A typical final state thus can contain either highly boosted top quarks, Higgs and gauge bosons, or a larger multiplicity of these particles with only moderate boost. The term "boost" refers to a particle with large kinetic energy compared to its rest mass, at hadron colliders defined using transverse momentum $p_\perp \gtrsim 2m$; the highly boosted regime roughly refers to particles from $p_\perp \gtrsim 5m$. When boosted particles decay, their decay products are collimated in the detector, leaving a characteristic image. Substantial progress has been made in the field of identification and reconstruction of boosted Standard Model resonances during recent years, see e.g. Refs. [18]–[21] for reviews. Many different proposals of tagging algorithms that do these tasks have been proposed and already included into experimental analyses. Most of them rely on the relatively new idea to collect all radiation from the resonance in one large-radius "fat" jet and study its substructure [22]. However, not quite as much focus has yet been put on tagging tops, Higgses, gauge bosons etc. at the very energy frontier to be probed soon.

When the final-state multiplicity is large and the fat jet intended to gather all radiation from the boosted resonance is not isolated from the rest of the event, most algorithms must struggle. Energy deposits may fall outside the jet if it is built too small, and radiation not correlated with the resonance may end up inside the jet if it is too large. In this thesis, we develop and discuss a novel jet algorithm that solves this problem in a unique way based on work by the author [A]. Dubbed "mass-jump clustering", it harnesses the strengths of fat jet substructure analyses [23], [24] without constructing the intermediate fat jet at all. We investigate the algorithm in different scenarios ranging from QCD dijets, via boosted top quarks, to very busy final states emerging from the production of TeV-scale BSM particles. For "standard" top tagging we find interesting properties and improved rejection of non-top-initiated jets. We develop a specific analysis for a benchmark BSM process with two boosted top quarks as well as two boosted Higgs bosons. We find that we can tag them with good quality, thus enhancing the regime of fat-jet based algorithms to very busy final states [B].

Also, highly boosted resonances force us to reconsider building jets from calorimeter information. The electromagnetic and hadronic calorimeters offer precise measurement of the energy deposit, but lack the spatial resolution when it comes to top quarks with boosts from $p_\perp \sim 1\,\text{TeV}$ or electroweak gauge bosons above $p_\perp \sim 500\,\text{GeV}$. We introduce dedicated taggers for these highly boosted tops and electroweak gauge bosons in this thesis based on work by the author [C]. These algorithms combine calorimetric

information with the good spatial resolution of the charged particle tracker, allowing good performance up to boosts of several TeV. This idea is not completely original and in fact we build on earlier work [24], [25], improving and enhancing performance as well as the field of applications. The energy regime under consideration will already be accessible at the LHC, but is also of importance at a future 100 TeV collider, both during design and analysis.

The core of this thesis is concerned with the tagging of top quarks, Higgs and electroweak gauge bosons at the very energy frontier. Boosted resonances have already been observed, and will be produced copiously during the upcoming run at the LHC, whether from Standard Model processes or from new physics on top of that. We improve on existing methods in two directions: multiplicity and boost. The thesis is outlined as follows. In Chapter 2 we briefly review the Standard Model and the role the top quark plays regarding electroweak symmetry breaking and new physics. We describe common jet clustering algorithms, top taggers and recent approaches based on fat jet shapes and clustering history in Chapter 3. From Chapter 4 we depict original work, and develop the mass-jump jet clustering algorithm and investigate its properties in conventional environments. In Chapter 5 we then show shortcomings of fat-jet-based analyses in very busy final states at the energy frontier, and apply the algorithm in a realistic study. Lastly we turn to highly boosted resonances and develop and investigate the `HPTTopTagger`, `HPTWTagger`, and `HPTZTagger` algorithms in Chapter 6. We review the key aspects of this thesis and draw conclusions in Chapter 7.

— That rug really tied the room together, did it not?

Walter Sobchak

2

The top quark in elementary particle physics

2.1 The standard model of particle physics

The Standard Model of Particle Physics (SM) is a relativistic quantum field theory that describes all known fundamental particles and their interactions (except gravity) in Nature. After the discovery of the Higgs boson in 2012 at the LHC [13], [14], its entire particle contents is now experimentally confirmed. It is based on the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The electroweak force emerges via the $SU(2)_L \times U(1)_Y$ group of Glashow-Weinberg-Salam theory [1]–[3], which is spontaneously broken by the Higgs field [9]–[12]. The $SU(3)_C$ symmetry describes the strong interaction between quarks and gluons known as quantum chromodynamics [4]–[7]. For a detailed discussion of the Standard Model we refer to the many textbooks (see e.g. [26]–[29]).

2.1.1 Electroweak theory

The electroweak sector of the Standard Model unifies the electromagnetic interaction via photons with the weak nuclear force, mediated by massive $W^{\pm}$ and Z gauge bosons. The matter content of the Standard Model is grouped into three copies of fermions with same quantum numbers, known as generations and summarized in Table 2.1. All fermionic fields carry $U(1)_Y$ hypercharge Y. The left-handed degrees of freedom transform as doublets under the group $SU(2)_L$ and carry weak isospin $T^3 = \pm 1/2$, while the right-handed fields are uncharged and transform as singlets. Electric charge Q is recovered by

$$Q = T^3 + \frac{Y}{2}\,. \tag{2.1}$$

				$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	Q
quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	3	2	$+1/3$	$+2/3$ $-1/3$
	u_R	c_R	t_R	$\bar{3}$	1	$+4/3$	$+2/3$
	d_R	s_R	b_R	$\bar{3}$	1	$-2/3$	$-1/3$
leptons	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	1	2	-1	0 -1
	e_R^-	μ_R^-	τ_R^-	1	1	-2	-1

Table 2.1: Transformation behaviour and quantum numbers of the Standard Model matter content.

When these symmetries are promoted to local (i.e. spacetime-dependent) symmetries, four massless gauge fields arise, W_μ^i and B_μ. A priori all fields in the Lagrangian are massless. The naïve introduction of mass terms

$$-m_f \bar{\Psi}\Psi = -m_f(\bar{\Psi}_L \Psi_R + \bar{\Psi}_R \Psi_L)\,, \tag{2.2}$$

$$\frac{1}{2} M_V^2 V_\mu V^\mu \tag{2.3}$$

for fermions and gauge bosons, respectively, is not possible as these terms explicitly break $SU(2)_L$ gauge invariance. In the Standard Model, the masses of all elementary particles are generated by the Higgs mechanism introducing a new elementary field Φ. It is an $SU(2)_L$ doublet of complex scalar fields with a quartic potential that develops a minimum at $\langle \Phi^\dagger \Phi \rangle = v^2$. This vacuum expectation value spontaneously breaks $SU(2)_L$ symmetry. Only one physical degree of freedom, the Higgs field H, remains and the other three degrees of freedom are "eaten" by the gauge bosons. Their mass eigenstates are the physical weak force particles $W_\mu^\pm$, Z_μ and the massless photon A_μ.

Yukawa couplings of the Higgs to the elementary fermions are invariant under $SU(2)_L$ and yield the desired mass terms after spontaneous symmetry breaking. As a result, all fermions interact with the Higgs field where the coupling constant is proportional to the respective fermion's mass.

2.1.2 Quantum chromodynamics

Quantum chromodynamics (QCD) is implemented as an $SU(3)_C$ gauge symmetry. Among the Standard Model fermions, only quarks are charged under this symmetry and come in triplets of three "colours". The corresponding gauge interaction is given by eight gluon fields G_μ^a, which mediate the strong nuclear force.

QCD phenomenology is shaped by an effect known as asymptotic freedom, which is understood from the perturbative expasion in the coupling constant α_S. In general, higher-order corrections in a quantum field theory exhibit ultraviolet divergencies. They can be handled by renormalization, a procedure that defines appropriate physical quantities from the bare quantities that appear in the Lagrangian. The renormalized coupling constant then incorporates higher-order effects and exhibits a dependence of the dimensionful renormalization scale, $\alpha \equiv \alpha(\mu^2)$. As physically meaningful results should not depend on

the arbitrary scale μ, one can derive a renormalization group equation for the coupling,

$$\frac{\partial}{\partial(\ln \mu^2)}\alpha(\mu^2) = \beta(\alpha(\mu^2))\,. \tag{2.4}$$

With the β function at one-loop order, the solution for the running coupling of QCD becomes

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + \frac{\alpha_S(\mu^2)}{12\pi}(33 - 2n_f)\log(Q^2/\mu^2)}\,. \tag{2.5}$$

Q^2 denotes the scale of the momentum transfer in the interaction vertex, and $n_f = 6$ is the number of different quark flavours. At large energy scales, the strong coupling constant gets increasingly small and quarks and gluons behave as asymptotically free particles. This is the energy regime of the hard interaction and the parton shower at collider experiments.

At small energy scales (corresponding to large distance scales), on the other hand, α_S increases and perturbation theory breaks down. As a result only composite $SU(3)_C$ singlet states (baryons and mesons, commonly denoted hadrons) are experimentally observed. This effect is known as confiment.

In a typical event at hadron colliders, the hard interaction of asymptotically free partons (quarks and gluons) is followed by a cascade of final-state radiation off the outgoing coloured particles. After the energy scale has evolved down to the regime governed by confinement, these "jets" of particles undergo hadronization. The detectors then measure the final state consisting of hadrons, charged leptons, and photons. Other sources of radiation are jets from initial-state radiation, underlying event (the soft reaction of the two beam remnants) and pile-up (more than one interaction in the same bunch crossing).

2.2 The special role of the top quark

The top quark stands out of the Standard Model's particle zoo in several ways. With a mass of roughly $m_t \approx 175\,\text{GeV}$ it is almost two orders of magnitude heavier than the second heaviest fermion, the bottom quark. In fact it is the only quark with a mass around the electroweak scale populated by the H, W and Z bosons, see Fig. 2.1. This hierarchy is not explained by the Standard Model, but has important implications on phenomenology.

After the discovery of the Higgs boson, the prime goal of the LHC is to establish the mechanism of electroweak symmetry breaking. Because the Yukawa couplings to the Higgs are proportional to the respective fermions' masses, the top quark exhibits a large coupling constant of $\mathcal{O}(1)$, connecting it tightly to the Higgs sector. In fact, Higgs production at the LHC is dominated by gluon fusion via a top quark loop and strongly depends on the t–H coupling [31]–[33]. A different process which has raised much attention is the production of a pair of top quarks in association with a Higgs boson (for some recent studies see Refs. [34]–[40], or Refs. [41]–[48] for experimental searches). This makes the top quark the key to Higgs physics, and to whatever lies beyond this mysterious elementary scalar (see illustration in Fig. 2.2).

The identification and reconstruction of top quarks with good efficiency is necessary for Standard Model physics and to pin down the mechanism of electroweak symmetry breaking. The tagging procedure is quite unique as the top is the only quark with a lifetime so short that it decays before hadronization [50].

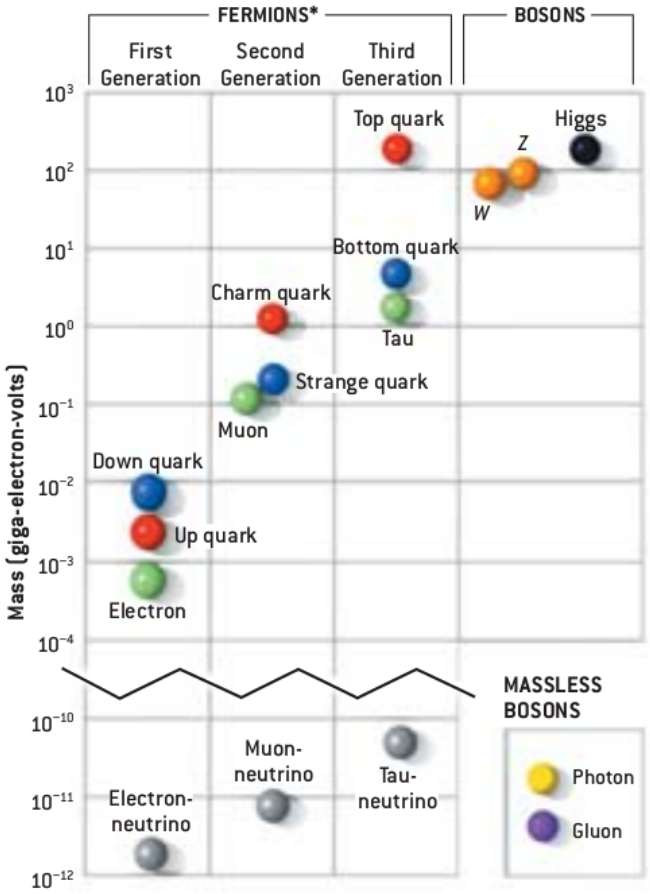


Figure 2.1: Masses of Standard Model particles (from Ref. [30]).

With more than 90% probability the decay proceeds in two steps starting from $t \to W^{\pm}b$. The W boson then further decays into a pair of leptons or quarks. This distinguishes the top quark also from the weak gauge bosons and the Higgs, which dominantly decay into only two prongs.

2.3 A key to new discoveries

During countless experimental tests at particle colliders during the last decades, the Standard Model has been crowned with success. With the discovery of the Higgs boson its particle contents is complete, which raises the question whether there might be no new discoveries at all up to the Planck scale. On the other hand we know that the Standard Model can not be the final answer. In particular when looking beyond collider experiments, there are a few obvious shortcomings that demand a substantial extension. As a first example, oscillations between different neutrino flavours have been observed, a fact that can only be theoretically explained if neutrinos carry a finite mass. Moving to the cosmological scale, there is

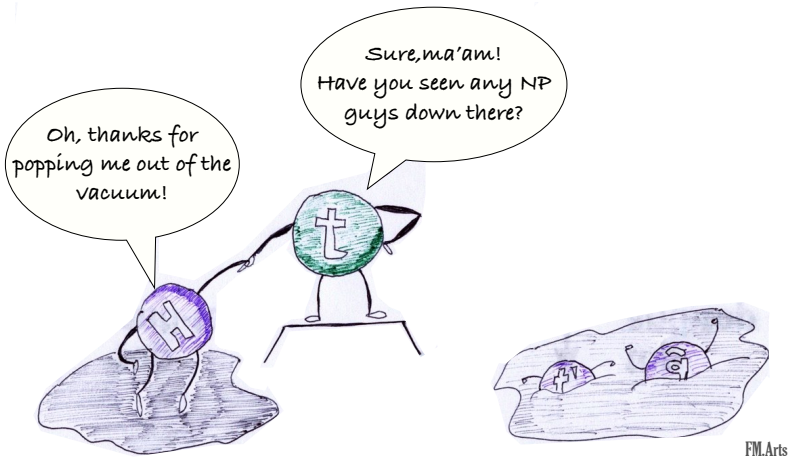


Figure 2.2: The top quark as a key to Higgs physics – opening the door to new physics (from Ref. [49]).

evidence for an unknown non-baryonic kind of matter, called Dark Matter. A suitable explanation could be given by a weakly-interacting massive particle (WIMP). There is no such candiate in the Standard Model, however. Other observations imply the existence of a new source of energy that leads to an accelerating universe, the Dark Energy. A cosmological constant as originally introduced in the Einstein equation of gravity can serve as a candidate, but its origin lacks an explanation. On a related note, there is still no convincing embedding of gravitation into quantum field theory. Furthermore, in order to explain the dominance of matter over antimatter in the universe, a process called baryogenesis (possibly by leptogenesis) is necessary to produce an asymmetry in the early stages after the Big Bang. The Standard Model cannot explain the observed asymmetry.

Besides these obvious shortcomings, there are also theoretical considerations that have guided the development of extensions to the Standard Model, and experimental searches. The hierarchy problem is introduced with the elementary scalar particle, the Higgs boson. Quantum corrections to its mass m_H are sensitive to physics at very large energy scales, and it is conceptually not clear why m_H is so much smaller than the Planck scale in the first place. To comprehend the problem, consider the one-loop correction to the Higgs mass Δm_H^2 induced by a massive Dirac fermion. If the ultraviolet divergence is regularized by a cut-off scale Λ_{UV}, the leading contribution reads

$$\Delta m_H^2 = -\frac{\lambda_f^2}{8\pi^2}\left[\Lambda_{\mathrm{UV}}^2 + \cdots\right] . \tag{2.6}$$

λ_f is the Yukawa coupling between the fermion and the Higgs boson, so the largest contribution stems from the heavy top quark. The Standard Model is known to be a renormalizable theory, i.e. the divergent terms can be absorbed into counterterms at any order in the perturbative expansion. The hierarchy problem (or similarly the fine-tuning problem, as the counterterm needs to be delicately adjusted in order to yield the physical Higgs boson mass), however, gives a hint as to where to expect new physics when the Standard Model is interpreted as an effective theory where resonances above the cutoff scale

Λ_{UV} have been integrated out. This scale can be estimated by the argument of naturalness. Here, the theory is called natural if the loop corrections to the Higgs mass are of the same order as the mass itself. In this sense BSM physics is expected at the TeV scale, and the LHC will be pushing into this regime during its upcoming run II.

Different models have been proposed in order to solve the hierarchy problem, out of which in the following two possible solutions are briefly introduced with a focus on new top quark production channels.

- **Composite Higgs**
  If the Higgs mass is protected against large mass corrections by a symmetry, the fine-tuning problem can be circumvented. First we briefly review a special class of composite Higgs models, called the Little Higgs models [51]–[54]. Little Higgs models introduce an enlarged approximate symmetry group, which is then spontaneously broken. The coset can be as simple as $[SU(3)/SU(2)]^2$ in the "simplest little Higgs" [55]–[57]. One of the pseudo-Nambu-Goldstone bosons that appear in the process is identified with the physical Higgs, which is now naturally much lighter than the scale of the underlying new strong dynamics. The explicit breaking (which is necessary to obtain a non-zero Higgs mass) of the symmetry is done "collectively", i.e. the symmetry is only broken when two or more couplings in the Lagrangian are non-vanishing. The Yukawa couplings to fermions are also embedded in the enlarged symmetry, which introduces a top partner that is necessary to cut off the quadratically divergent loop in the Higgs mass, Eq. (2.6). This heavy particle is called the vectorlike top T and can be directly searched at particle colliders from its decay channels $T \to tH$, $T \to tZ$, and $T \to bW^+$ (some models also contain a bottom partner B or more exotic states $T_{5/3}$ and $B_{-4/3}$). By naturalness arguments, its mass is expected to be $\lesssim 2\,\mathrm{TeV}$ [16]. Heavy $SU(2)$ gauge bosons are predicted and can be readily searched in collider experiments, while the production rate and partial decay widths are model-dependent. In order to recover hypercharge, an additional $U(1)_X$ is gauged, and the final particle spectrum also contains a heavy gauge boson called Z'. Other cosets discussed in the literature are $[SU(3)_L \times SU(3)_R/SU(3)_V]^4$ (Minimal Moose [58]), $SU(5)/SO(5)$ (Littlest Higgs [59], [60]), and many more (for a review, see e.g. Ref. [16]).

  Technicolour models introduce an additional strong sector [61], [62] (for reviews also see Refs. [63]–[65]). Similar to QCD, the theory is asymptotically free but becomes confining at the electroweak scale. The global chiral symmetry of new, massless fermions is spontaneously broken by the formation of a "technifermion" condensate. Through this process, Standard Model gauge bosons and fermions can aquire mass. The large mass of the top quark compared to the other fermions in the Standard Model suggests that the third generation plays a special role in the process of electroweak symmetry breaking. Embedding the $SU(3)_C$ gauge group of QCD into a larger structure, e.g. $SU(3)_1 \times SU(3)_2$ where $SU(3)_2$ ($SU(3)_1$) only couples to the third (first and second) generation, leads to topcolour models. The top quark condensate, a $t\bar{t}$ bound state, is identified with a composite Higgs boson. Together with technicolour (or a different mechanism that can provide the correct scale of the top mass) these models can give the correct Higgs mass and allot a special role to the third generation of quarks. One more component is necessary to enhance the formation of the $t\bar{t}$ condensate, while suppressing the formation of the $b\bar{b}$ condensate in order to obtain a relatively light b quark. This so-called tilting can be achieved by an additional neutral gauge boson

Z'. In Refs. [66], [67] a leptophobic, topophyllic topcolour Z' is introduced that dominantly decays into a pair of top quarks.

- **Supersymmetry (SUSY)**
  To see how supersymmetric models cure the hierarchy problem, consider the one-loop correction to the Higgs mass from a complex scalar field,

$$\Delta m_H^2 = \frac{\lambda_s}{16\pi^2} \left[\Lambda_{\mathrm{UV}}^2 + \cdots\right] . \tag{2.7}$$

  Comparing this expression to Eq. (2.6), it is observed that the quadratically divergent terms cancel exactly if the number of fermionic and bosonic degrees of freedom is identical and the respective couplings fulfill $\lambda_s = \lambda_f^2$. For the leading contributions from massive Standard Model particles, this implies the existence of two scalar top partner fields as well as fermionic partners for the electroweak gauge bosons. Supersymmetry (see e.g. Ref. [15]) introduces such superpartners for each elementary particle, thus solving the hierarchy problem. To prevent couplings that violate lepton and baryon number conservation and mediate the SUSY-induced decay of the proton, R-parity conservation is typically imposed. Superpartners can then only be produced in pairs, and the lightest supersymmetric particle (LSP) is stable and can serve as a dark matter candidate.

  Supersymmetry also gains theoretical motivation from the Coleman-Mandula no-go theorem [68]. This theorem states that there is no interacting quantum field theory where the generators of the internal symmetries of the S-matrix are combined with the Poincaré group in any but a trivial way. However, supersymmetry introduces antisymmetric Fermi-type generators that avoid these constraints. It has been shown [69] that the supersymmetric Super-Poincaré group is the only graded Lie algebra that leads to a consistent quantum field theory, ergo supersymmetric theories possess the largest possible spacetime symmetry. On a last note, the running gauge couplings can be unified at high energies $\sim 10^{16}\,\mathrm{GeV}$, a feature that is not possible in the Standard Model.

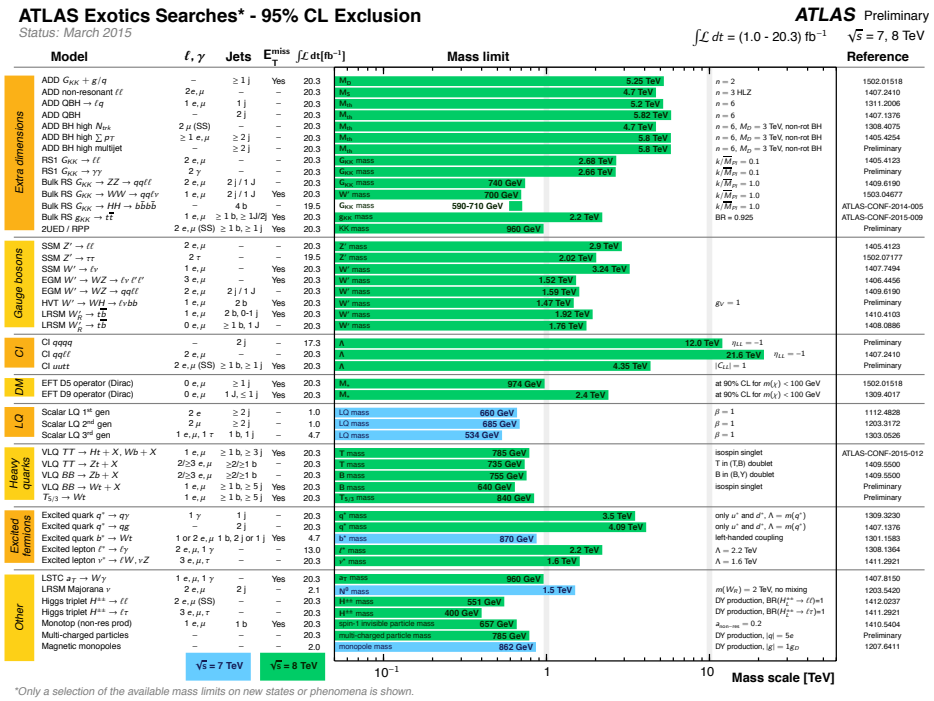
  Already the Minimal Supersymmetric Standard Model (MSSM) introduces more than 100 parameters in addition to the Standard Model parameters, motivating simplified models under additional assumptions for phenomenological analyses. One of the simplest models, the Constrained MSSM, however, could already be excluded at 95% confidence level by combined results [70]. In particular all experiments to date gave null results in direct searches of the predicted superpartners, which suggests that supersymmetry is broken and the superpartners potentially carry a large mass. A typical decay chain of a heavy superpartner contains boosted Standard Model particles and missing energy from the undetected LSP.

  Among other non-minimal setups, vectorlike top partners can be added to supersymmetric models to increase the light Higgs boson mass while other SUSY particles are kept relatively light [71]–[73]. This can simulateously explain the measured Higgs boson mass and the discrepancy of the muon anomalous magnetic moment with its Standard Model prediction [74]–[76].

There are many more theoretical concepts that predict heavy resonances, some of which decay into top quarks. So far none of these predicted particles could be observed experimentally, and Fig. 2.3 shows exclusion bounds from ATLAS and CMS after run I of the LHC. The experimental searches referenced

there can also serve as a good starting point to study relevant BSM models that are currently under investigation.

A novel search for the vectorlike top with subsequent decay $T \to tH$ is developed in Chapter 5 based on a global top quark and Higgs boson tagging algorithm. Chapter 6 describes the search for a very heavy topophyllic topcolour Z' decaying into a pair of highly boosted top quarks using the new `HPTTopTagger`. The tagger is also adapted to identify highly boosted W and Z bosons in searches for a heavy charged gauge boson W'.



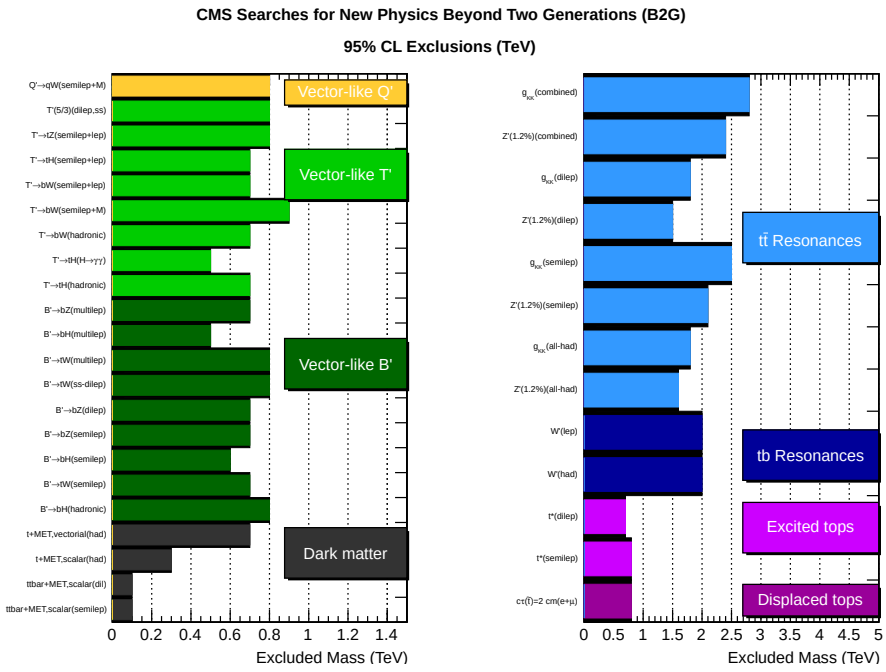


Figure 2.3: Exclusion bounds on heavy resonances as observed by ATLAS (upper, from Ref. [77]) and CMS (lower, from Ref. [78]) during LHC run I.

— Never send a human to do a machine's job.

Agent Smith

3 Top quark tagging

Now that the motivation or the "why" has been established, we turn to the "how". Top quark tagging describes the procedure of identification and possibly also kinematic reconstruction of top quarks, and the simultaneous rejection of backgrounds of different origin. From the first discovery 20 years ago, the focus recently shifted towards reconstruction of hadronically decaying top quarks with large transverse momentum. This "boosted" regime has been made accessible by the large centre-of-mass energy of the LHC on the one hand, as well as progress in jet substructure analysis methods on the other hand. In this chapter, we give a brief review of the history of top quark tagging and finish with an overview of some of the essential state-of-the-art tagging algorithms. On this journey we go into the details of sequential jet clustering, point out sources that disturb jet–parton correspondance at the LHC as well as possible solutions, then describe intrinsic jet properties known as shapes, before finishing off with current taggers based on jet clustering history. The main parts of this thesis are later built on the firm foundations set in this very chapter: a new class of jet algorithms in Chapter 4 and its application to very busy final states in Chapter 5, as well as dedicated tagging algorithms for highly boosted top quarks and electroweak gauge bosons based on jet substructure in Chapter 6. The high-energy frontier is penetrated from two directions, high-multiplicity and high-energy.

3.1 Overview

At collider experiments, not the final-state particles of a hard process are detected, but rather their trace in the detector in the form of charged tracks, electromagnetic and hadronic calorimeter energy deposits ("towers"), and muon signals. When pseudo-events are simulated with Monte Carlo event generators,

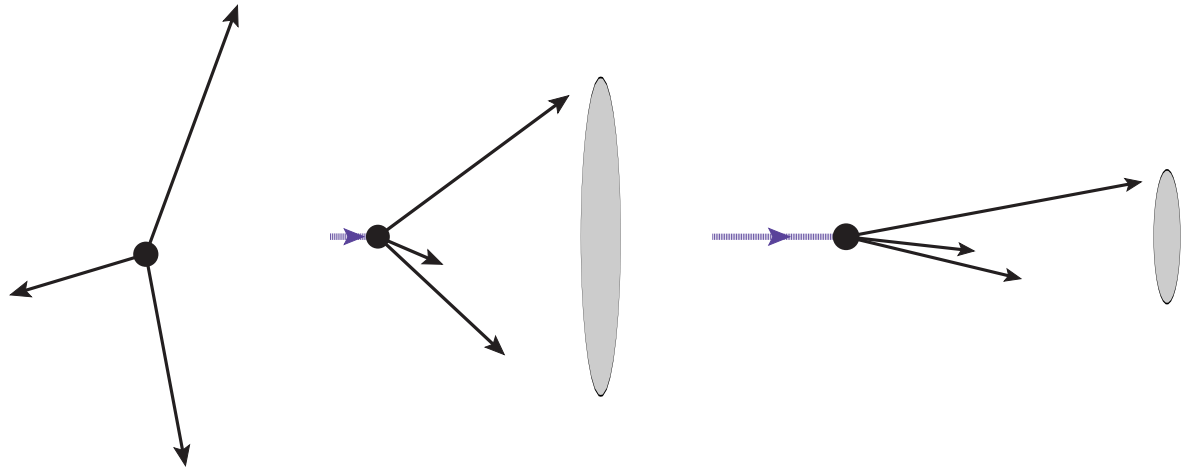


Figure 3.1: Visualization of the top decay $t \to bW^+ \to bjj$ in different frames of observation. In the top quark rest frame, the decay products are well-separated (left). When the top quark is boosted, the three prongs start to merge and can be caught in a single large-radius "fat" jet (middle). In the highly boosted regime, a smaller fat jet is sufficient to catch all radiation. Angular distances between the prongs may become too small to be resolved by the calorimeter (right).

one distiguishes between three stages of the final state: The parton level, which is the evolved final state after parton showering consisting of the elementary quarks, gluons, leptons, etc. – the hadron level, which is the state after hadronization and hadronic decays and is supposed to be the realistic set of particles entering the detector – and finally the detector level, which is the simulated detector response, including deficiencies and finite resolution effects.

Isolated charged leptons ($e^{\pm}$, $\mu^{\pm}$) can be identified by their clear signal, and the undetected neutrinos show up in the form of missing energy. Coloured particles on the other hand are only reconstructed as finitely distributed jets from the energy deposits in an approximately circular neighbourhood of the detector (usually characterised by an angular radius parameter R). The jets are assumed to correspond to the final-state light quarks and gluons after parton shower and hadronization. Assuring this kinematic one-to-one correspondance is the task of jet algorithms. A splendour of jet reconstruction algorithms has been proposed, and starting from the next section we go into many more details about jet properties and analysis methods.

It was the year 1995 when the CDF [79] and DØ [80] collaborations simulateously announced the discovery of the top quark at the Tevatron proton-antiproton collider. This completed the matter content of the Standard Model, and it would take another 17 years to find the last (presumably) elementary particle, the Higgs boson. Both collaborations found an excess in events with one or two isolated charged leptons, missing transverse energy, and additional jets (including bottom quark tags). In the channel where one charged lepton and four (or more) jets could be identified, they were kinematically fitted to the assumed final state $t\bar{t} \to W^+W^-b\bar{b} \to l\nu q\bar{q}'b\bar{b}$, and the reconstructed top quark mass could be extracted.

At Tevatron centre-of-mass energy of $\sqrt{s} = 1.8\,\text{TeV}$, the top quarks do not acquire large momentum and their decay products are well separated in the detector. Still at the LHC the majority of top quarks are produced with low momentum. However, due to the much larger collision energy, boosted tops became accessible and will be copiously produced during run II and even more at a future 100 TeV collider. The term "boosted" was coined for heavy particles whose transverse momentum exceeds twice their rest mass, $p_\perp \gtrsim 2m$. Phenomenology significantly changes, as the decay products are now collimated in the laboratory frame. (Also see Fig. 3.1.) For the 1–to–2 decay of a heavy particle with mass m and transverse momentum $p_\perp$, the separation is typically $\Delta R \gtrsim 2m/p_\perp$.[1] At high boosts, the individual jets become very close, merge and cannot be separated with classical methods any more. If instead all decay products are caught in one large-radius jet, its multi-prong substructure differs characteristically from pure single-pronged QCD jets. This proposal [22] led to a highly active field of research in recent years focusing on those "fat jets". Fig. 3.2 schematically depicts different regimes of top boosts and event multiplicity and a selection of tagging algorithms. The remainder of this chapter briefly recaps these algorithms. Here and throughout this thesis, we focus on the hadronic decay of the top quark $t \to bW^+ \to bjj$. The reason is that the classic handles of the semileptonic decay – a bottom-flavoured jet, an isolated charged lepton, and large missing transverse energy $E_\perp^{\text{miss}}$ from the undetected neutrino – are diminished when the top quark is boosted. First of all, bottom tagging discrimination power is degraded at high $p_\perp$ [81]. Then, as the decay products of boosted top quarks tend to be collimated, the charged lepton often merges with the b jet and isolation criteria fail.[2] The neutrino can have low transverse momentum and not leave a clear trace of $E_\perp^{\text{miss}}$. Therefore the hadronic decay is more promising for boosted top quarks, due to its larger branching fraction $\sim 2/3$.

To channel the developments in tagging of boosted resonances, the BOOST conference was established and is held every year since 2009. The conference reports are an excellent source of recent progress and benchmarks. However, a trend can be seen when comparing the performance of various top tagging algorithms in the reports of the 2010 [18] and 2011 conferences [19]: Progress has significantly slowed down and many algorithms show remarkably similar performance, cf. Fig. 3.3. The algorithms included in the comparison are listed in Tab. 3.1. In the 2012 summary [20], the focus already shifted to experimental validation and the study of jet substructure from first principles. The latest report to date, summarizing BOOST2013 [21], also does not add any new top tagging algorithm with superiour performance to the analysis portfolio, but rather focuses on correlations to improve performance. This suggests that based on the ideas and data currently on the market, there is only little room left for further improvement.[3] Consequently the goal of this thesis is to fill those gaps in Fig. 3.2 that become relevant at LHC run II.

[1] At hadron colliders, the angular separation between two particles is defined $\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2}$ where Δy and $\Delta\phi$ are the distances in rapidity and azimuthal angle, respectively.

[2] For studies of semileptonically decaying boosted tops see e.g. Refs. [82]–[84].

[3] A combination of orthogonal observables can significantly improve jet discrimination compared to the respective stand-alone algorithms, see e.g. Refs. [85] and [21].

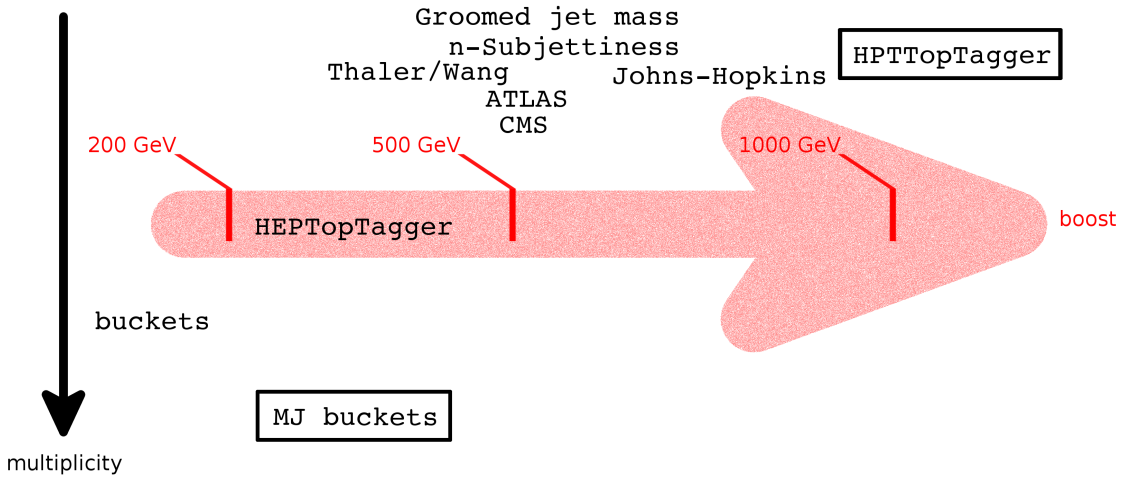


Figure 3.2: Schematic overview of selected top tagging algorithms and their favoured application regime in terms of boost (horizontal axis) and multiplicity in the event (vertical axis). Most algorithms operate in the low-multiplicity region, in other words the top quark is required to be sufficiently separated from other hard radiation. The algorithms marked with bounding boxes cover two regions that become accessible during run II at the LHC and are developed in this thesis: mass-jump jet clustering in combination with buckets in Chapter 5, and the `HPTTopTagger` in Chapter 6. The other algorithms are referenced in Tab. 3.1.

- `ATLAS default top tagger` [86], [87], reviewed in Sec. 3.6.
- `CMS top tagger` [88], reviewed in Sec. 3.6.
- `HEPTopTagger` [24], reviewed in Sec. 3.6.
- `Johns Hopkins tagger` [89], reviewed in Sec. 3.6.
- N-Subjettiness [90], [91], reviewed in Sec. 3.4.
- Pruned jet mass [92], [93], reviewed in Sec. 3.3.
- `Thaler/Wang top tagger` [82], we refer to the literature.
- Trimmed jet mass [94], reviewed in Sec. 3.3.

Table 3.1: Selection of common top tagging algorithms with reference to their reviews.

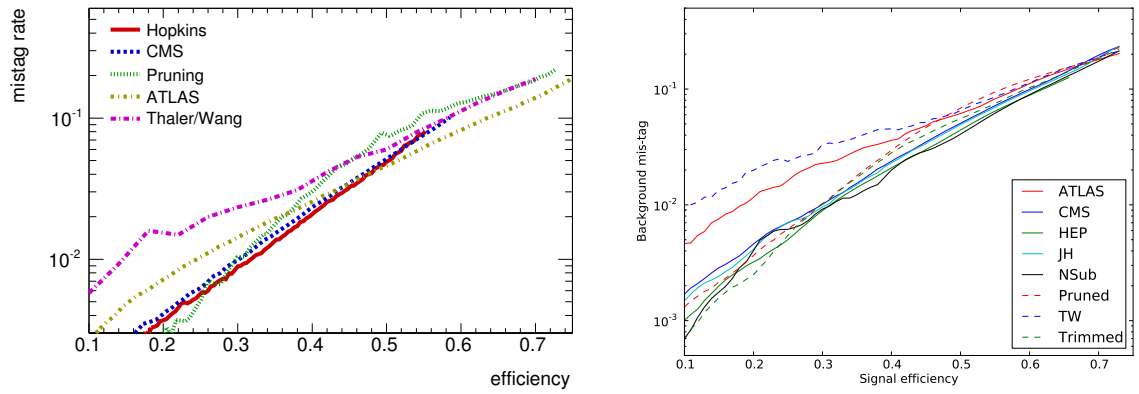


Figure 3.3: Receiver-operator characteristic (ROC) curves for several top tagging algorithms, taken from the BOOST2010 (left) and BOOST2011 (right) conference reports. The horizontal axis depicts the efficiency to correctly identify a boosted top quark, and the vertical axis depicts the corresponding mis-identification rate for QCD-initiated jets. Events were generated with the HERWIG 6.510 [95] event generator for the LHC at $\sqrt{s} = 7\,\text{TeV}$. The efficiencies are calculated from a flat distribution of fat jets in $p_\perp$, ranging from 200 GeV to 1.6 TeV. The different algorithms are listed in Tab. 3.1.

3.2 Jets at the LHC

The standard jet clustering algorithms used at the LHC are members of the infrared and collinearly safe family of generalized-k_T sequential recombination algorithms. Input to these algorithms are so-called topo-clusters, objects constructed from long-lived particles' energy deposits in the electromagnetic and hadronic calorimeter [96]. The CMS collaboration primarily uses so-called particle flow objects [97], which are constructed from all visible particles in an event, i.e. muons, electrons, photons, charged hadrons, and neutral hadrons. The topo-clusters or alternatively particle-flow objects (or pseudo-particles in simulated data) are referred to as protojets in the context of jet clustering. An angular distance measure is defined between pairs of them,

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min\left[p_{i\perp}^{2n}, p_{j\perp}^{2n}\right] , \quad d_{iB} = p_{i\perp}^{2n} , \tag{3.1}$$

where $n = 1$ corresponds to the k_T algorithm [98]–[100], $n = 0$ to the Cambridge/Aachen algorithm [101], [102], and $n = -1$ to the anti-k_T algorithm [103].

The algorithms then proceed as follows:

1. Find the smallest d_{ij} among protojets. If it is given by a beam distance, d_{iB}, label i a jet and repeat step 1.

2. Otherwise combine $ij \to k$ by summing their four-momenta, $p_k = p_i + p_j$ (E-scheme, see e.g. Ref. [104]). In the set of protojets, replace i and j by k and go back to step 1.

Clustering eventually terminates when all particles have been merged into jets. All jets are separated by

an angular distance larger than the radius parameter R, because a minimal beam distance d_{iB} in the clustering sequence implies $\Delta R_{in} > R \; \forall$ protojets n.

Whenever jets are used as input for an analysis, the significance of the results crucially depends on the validity of the kinematic correspondance to the respective jet-initiating partons. Besides the before-mentioned sequential recombination algorithms, there has been ongoing effort to construct new and improved jet algorithms that are also infrared and collinear safe. Some proceed via cones [104]–[107], while others follow completely different original ideas [108]–[110]. In the majority of these algorithms, jets are constructed with fixed angular size R, though.

Despite the large portfolio of jet algorithms to select from, choosing the optimal radius R is always a compromise [111]–[113] as it may be different for jets of different energy or position in the detector. Possible effects that compromise the kinematic validity of the reconstructed jet and that are directly related to the jet radius are splash-out (not all final-state radiation is caught inside the jet) and splash-in (radiation from a different source such as another nearby hard parton ends up inside the jet).

Hadronic final states are subject to other sources of QCD radiation as well. Soft and collinear jets from incoming partons are labelled initial-state radiation, and are well described by the parton shower. Underlying event denotes soft QCD activity that arises from interactions of the proton remnants. Its distribution is isotropic and depending on the jet radius it can have a large effect on the jet's invariant mass [111],

$$\langle \delta m_j^2 \rangle \simeq \Lambda_{\mathrm{UE}} \, p_{\perp,j} \left(\frac{R^4}{4} + \frac{R^8}{4608} + \mathcal{O}(R^{12}) \right) . \tag{3.2}$$

Λ_{UE} is the amount of the transverse momentum contribution from the underlying event per unit rapidity. At the LHC, $\Lambda_{\mathrm{UE}} \sim \mathcal{O}(10)\,\mathrm{GeV}$ [114]. Multiple proton-proton collisions in the same bunch crossing are known as pile-up. Its effects have already been observed and during run II at the LHC, pile-up levels are expected to peak around 40–50 [115], [116]. Similarly to initial-state radiation and underlying event, pile-up effects diminish the accuracy of the reconstructed jet momenta particularly for large radii and in a realistic setup they have to be removed with grooming methods. The essential algorithms are introduced in the context of jet mass in the following section.

3.3 Jet mass and grooming

The invariant mass of the four-momentum of a jet is known as (plain) jet mass. In the parton-shower picture, it is induced by perturbative QCD emissions off the jet-initiating coloured parton, with each emission accounting for a contribution

$$\delta m^2 = 2z(1-z)E^2(1-\cos\theta) \tag{3.3}$$

to the jet mass. The two daughter partons carry energy fractions zE and $(1-z)E$ of the mother parton, and the emission is described by the opening angle θ in an angular-ordered parton shower (other common approaches are parton showers ordered in virtuality or transverse momentum of the emissions). Under certain approximations jet mass has been studied analytically and differences between quark- and gluon-initiated jets as well as the impact of different jet definitions have been investigated [117]–[120] (for a qualitative discussion see Ref. [21]).

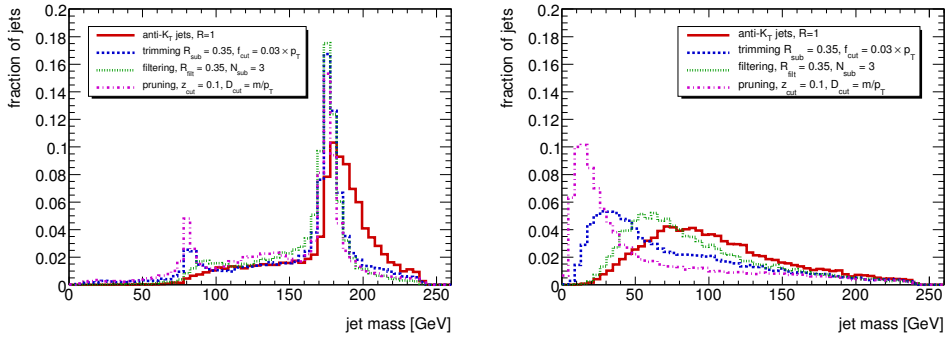


Figure 3.4: Jet invariant mass for $t\bar{t}$ (left) and QCD dijet (right) events with parton-level $p_\perp \in [500, 600]\,\text{GeV}$, taken from Ref. [18]. The red solid line depicts plain jet mass. The effects of the different grooming methods described in the text are visualized by the other three curves, with parameters given in the legend.

There are two main questions related to top quark (or equally gauge boson, Higgs, etc.) tagging. The first is the origin of jet mass, which is perturbatively induced as described above and relevant for a description of the QCD background. For a fat jet that contains all decay products of the top quark, the heavy mass is the result of the 4-momenta sum of its three reconstructed partonic constituents. The plain jet mass is therefore correlated with the top parton mass and allows discrimination against QCD jets, which typically have much lower values, see Fig. 3.4.

As discussed in Sec. 3.2, uncorrelated radiation stemming from underlying event or pile-up can increase the measured mass and thus impede tagging performance especially for the large-radius jets typically considered. Hence the second question is how to identify signal in such noisy environments.

The following three tools proved successful in "grooming" the jet from this soft radiation and can give a signifant improvement on the plain jet mass, in particular by pushing background jets to lower values. While they were originally proposed in different contexts, they all rely on the same observation: The energy of the uncontaminated jet is concentrated in only a few small regions, where on the other hand radiation from underlying event and pile-up has characteristically lower energy deposits. All groomers start from a jet with radius R and transverse momentum $p_{\perp,J} \geq p_\perp^{\min}$.

Trimming [94]: The constituents of the jet are reclustered into subjets of smaller radius $R_{\text{trim}} < R$. Only subjets i with transverse momentum $p_{\perp,i} \geq f_{\text{cut}}\, p_{\perp,J}$ are accepted and recombined into the trimmed jet. Ideally, QCD jets only consist of the leading subjet after trimming, while in fat jets from a heavy resonance decay, all subjets associated with the hard prongs remain. Typical parameters are $R_{\text{trim}} = 0.2 \sim 0.35$ and $f_{\text{cut}} = 0.03$.

Filtering [23]: Similarly to trimming, the jet is reclustered into subjets with radius R_{filt}. In most applications, the Cambridge-Aachen jet algorithm is applied. The filtered jet is then built from the n_{filt} hardest subjets, where n_{filt} is typically one or two more than the number of hard partons in the decay of the heavy resonance to also include the leading gluon emissions.

Pruning [92], [93]: Pruning is designed to discard large-angle and soft radiation during the sequential clustering of the jet. A recombination step is vetoed if the angular distance between the two protojets i, j is large, $\Delta R_{ij} > R_{\text{cut}} \frac{2m_J}{p_{\perp,J}} \equiv D_{\text{cut}}$, and the transverse momentum is asymmetric, $\frac{\min(p_{\perp,i}, p_{\perp,j})}{p_{\perp,ij}} < z_{\text{cut}}$. In this case, the softer of the two protojets is discarded. Otherwise clustering proceeds as usual. Common choices for the parameters are $z_{\text{cut}} = 0.1$ and $R_{\text{cut}} = 0.5$.

Fig. 3.4 shows the impact of the three groomers on fat jets in $t\bar{t}$ event and QCD dijets. For the chosen parameters pruning and trimming act most aggressive, resulting in low QCD jet masses and an additional peak around the W mass for t-initiated jets when one of the hard prongs is mistakenly groomed away. Conversely, the filtered jet always includes n_{filt} subjets and does not form a second peak in signal events. Likewise single-pronged QCD jets cannot be stripped to the hardest prong only such that the filtered mass is still relatively large.

Grooming methods are not the only answer to the questions above given in the literature. Jet shapes based on the energy flow inside a jet (see Sec. 3.4) and the identification of subjets (see Secs. 3.5 and 3.6) are currently under lively discussion, too.

3.4 Jet shapes

In order to discriminate between jets of different origin, say top-quark-initiated jets versus light-quark- or gluon-initiated jets, it is necessary to probe the inner structure of the jet. Jet shapes, which are often derived from classical event shapes [121], are substructure observables that take the four-momenta of the jet's constituents as their argument. Jet mass was discussed in the previous section and constitutes the simplest jet shape.

Another basic question one can ask is the number of hard subjets inside a given jet. N-subjettiness [90], [122] quantifies the degree to which the energy flow is aligned along N specified subjet axes. It is defined as

$$\tau_N^\beta = \frac{\sum_i p_{\perp,i} \min\left(\Delta R_{1i}^\beta, \Delta R_{2i}^\beta, \cdots, \Delta R_{Ni}^\beta\right)}{R_0^\beta \sum_i p_{\perp,i}}, \tag{3.4}$$

where the sums run over the jet constituents, R_0 is chosen such that $\tau_N^\beta < 1$ and the exponent β is typically set to one. Optimal axes are those that minimize N-subjettiness. Small values of τ_N^β indicate that the energy distribution inside the jet is localized around the N axes, i.e. it contains at most N relevant hard prongs. The ratio

$$\tau_{N,N-1}^\beta \equiv \tau_N^\beta / \tau_{N-1}^\beta \tag{3.5}$$

is a powerful discriminant in top quark (τ_{32}^β) and W boson (τ_{21}^β) tagging, and it is less prone to effects of typical QCD radiation than plain N-subjettiness. It can be made infrared-collinear safe with a loose lower cut on τ_{N-1} and used as a tagger.

Shower deconstruction [123], [124] is a highly effective approach, which calculates the probability that a parton shower MC simulation of a hadronically decaying top quark reproduces the jet under consideration including its substructure. The same calculation is repeated for all relevant background

processes, and from the two probabilities a likelihood ratio is formed. Because all possible radiation patterns need to be taken into consideration, the evaluation is computationally very expensive. Elaborate implementation and large computation time are the reasons why shower deconstruction is rarely applied, despite outperforming all other algorithms developed to date.[4]

Any algorithm that relies on subjets is limited by the possibility that a wrong set of subjets is chosen by the jet algorithm. Because sequential jet reconstruction is deterministic, one intermediate clustering step that does not correspond to the shower history can compromise the results. Qjets [125], [126] generalize the distance measure d_{ij} (Eq. (3.1)) to a probability

$$P_{ij} \propto \exp\left(-\alpha \frac{d_{ij} - d_{\min}}{d_{\min}}\right) , \tag{3.6}$$

according to which the recombination step is chosen randomly. The rigidity α is typically chosen 0.1 or 0.01, and $d_{\min}$ is the smallest distance measure in the respective recombination step. The classically inaccessible mass volatility

$$\mathcal{V} = \frac{\sqrt{\langle m^2 \rangle - \langle m \rangle^2}}{\langle m \rangle} \tag{3.7}$$

is calculated from the pruned jet mass (cf. Sec. 3.3) of a number of Qjets clustering histories and tends to be smaller for W jets compared to QCD jets. The same observable has also been applied in quark/gluon discrimination, see e.g. Ref. [21] for an overview.

Many more genuine jet shapes exist in the literature, and here we gave only a very small selection. For a brief overview in the context of boosted resonance tagging, the reader is referred to the BOOST reports [18]–[21].

3.5 Jet clustering history: the beginnings

The three canonical sequential jet clustering algorithms of the generalized-k_T family (cf. Sec. 3.2) all reconstruct jets with a mutual separation larger than the radius parameter R, $\Delta R(j_a, j_b) > R \quad \forall a \neq b$. Each recombination step $ij \to k$ involves protojets separated by less than the radius, $\Delta R(i,j) < R$. However, due to different distance measures, the order of recombination steps, known as clustering history, differs significantly.

The k_T algorithm was proposed first, and was designed to reverse the splitting history of the physical process, in particular the parton shower. Mergings of soft and collinear protojets, which are enhanced in the parton shower splitting kernels, are favoured. This is manifest in the distance metric

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min\left[p_{i\perp}^2, p_{j\perp}^2\right] , \quad d_{iB} = p_{i\perp}^2 . \tag{3.1 a}$$

When all decay products of a heavy resonance are caught in one fat jet, the final recombination step therefore usually corresponds to the merging of the two energetic decay products to the heavy object. The first W boson tagger [22] based on jet substructure exploited this fact by un-doing the last clustering

[4] It has been shown [85] that similar or even better performance can be achieved in a multivariate approach including the `HEPTopTagger`, N-subjettiness, and Qjets.

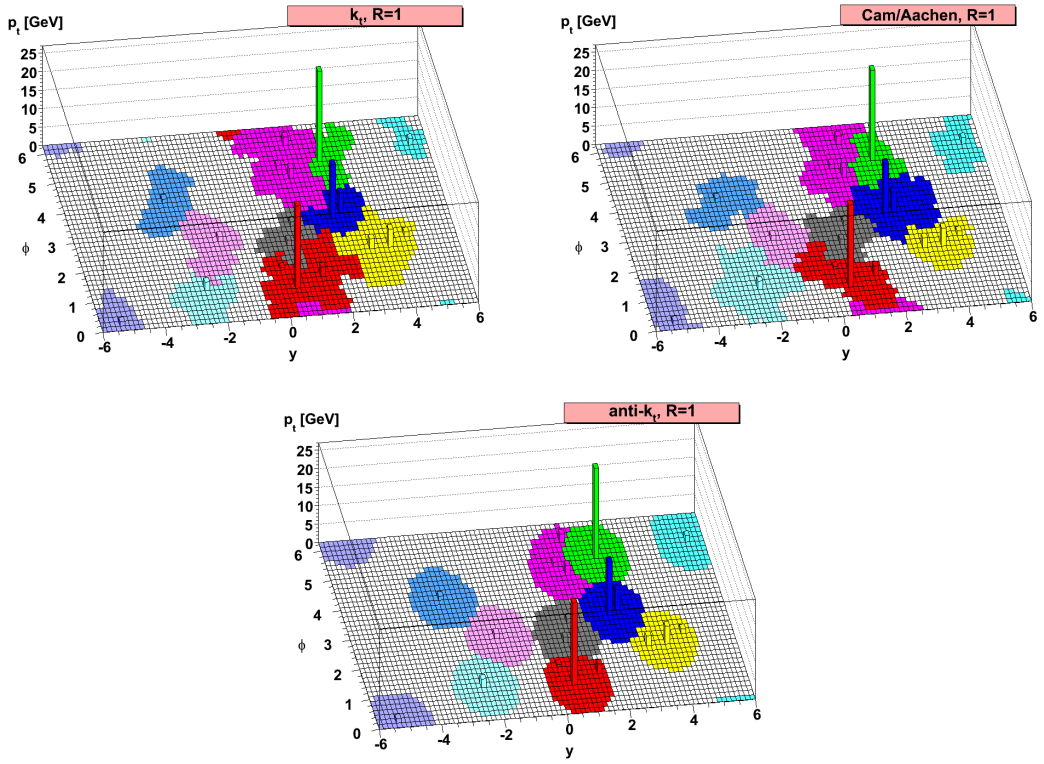


Figure 3.5: The "active catchment areas" in the η–ϕ plane of the final hard jets obtained with different jet clustering algorithms for the same event. The towers denote cells with energy deposit. The areas of each jet are obtained by including randomly distributed "ghost" particles with infinitesimal transverse momentum in the clustering procudure: those ghosts that end up in a certain jet define its area [128]. Figures taken from Ref. [103].

step to obtain two subjets. A cut on their R separation constitutes the tagging algorithm. Whereas QCD jets can also acquire a mass of order m_W, this typically happens through large-angle soft radiation.

A different algorithm, the so-called `YSplitter` [127] simply employs a cut on the value of $y_{1-2} \equiv d_{ij}$ in the final merging step. This value is correlated with the mass of the heavy resonance and QCD events tend to have very small y_{1-2}. A generalization of this approach to the two-step top decay directly leads to the `ATLAS default top tagger` and will be described in Sec. 3.6.

One characteristic of the k_T algorithm is that in the early clustering stage, only soft protojets are recombined. This leads to the formation of "junk" clusters, from which the final jets are built. As a result, k_T jets can experience quite different geometric shapes and areas even for the same radius R, see Fig. 3.5. This is to some extend avoided by the Cambridge-Aachen algorithm, which employs a purely

geometric distance metric,

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2}\,, \quad d_{iB} = 1\,. \tag{3.1 b}$$

The corresponding clustering history is in closer resemblance to an angular-ordered parton shower, and has been exploited in the `BDRS Higgs tagger` [23]. In contrast to the k_T algorithm, however, it is not sufficient to undo the last clustering step. Because of the absence of any momentum scale in the distance metric, soft and uncorrelated radiation from the edge of the jet is often clustered during the final steps. The `BDRS` algorithm works its way backwards through the clustering history until a substantial "mass drop" is found,

$$\frac{\max[m_i, m_j]}{m_{ij}} < \mu \quad , \quad \frac{\min[p_{i\perp}^2, p_{j\perp}^2] \cdot \Delta R_{ij}^2}{m_{ij}^2} \sim \frac{\min[p_{i\perp}, p_{j\perp}]}{\max[p_{i\perp}, p_{j\perp}]} > y_{\text{cut}}\,. \tag{3.8}$$

The second condition requires the energy sharing in the splitting to be symmetric. In any splitting where these condidtions are not met, the less massive prong is discarded.[5] The cut is governed by two dimensionless parameters μ and y_{cut} (with default values $\mu = 0.67$, $y_{\text{cut}} = 0.09$) and un-clustering can continue down to arbitrarily small scales, making the algorithm well-suited for searches of boosted resonances of unknown mass. If a mass drop is found, the two subjets are then filtered (cf. Sec. 3.3) with a radius

$$R_{\text{filt}} = \max(\ 0.2,\ \min(0.3, \Delta R_{j_1 j_2}/2)\ )\ , \tag{3.9}$$

and the Higgs momentum and mass are reconstructed from the three hardest filtered subjets to also capture the leading gluon emission. If no mass drop is found, the candidate is discarded.

The anti-k_T algorithm is the most recent of the generalized-k_T clustering algorithms, and favours the recombination of protojets with large transverse momentum.

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min\left[p_{i\perp}^{-2}, p_{j\perp}^{-2}\right]\,, \quad d_{iB} = p_{i\perp}^{-2}\,. \tag{3.1 c}$$

Jets are sequentially built around hard prongs, leading to circular jets in the y–ϕ plane (unless jets are not separated by $\Delta R(j_1, j_2) > 2R$ and overlap) as expected from an experimental point of view, see Fig. 3.5. Although the distance metric is formulated similarly to the k_T and C/A algorithms, its clustering history is generally not suited for substructure analyses. The reason is that when two hard prongs are merged in a fat jet, this does not happen in a single step but the hardest protojet typically absorbs the softer one across a multitude of recombination steps.

3.6 Top tagging with fat jet unclustering

Due to their large mass, boosted top quarks became accessible only after the electroweak gauge bosons. The tagging algorithms described below all build on the ideas based on the hadronic W or Higgs boson taggers outlined in Sec. 3.5, and are adapted to the two-step decay of the top quark.

[5] It has been pointed out in a related setup [119] that following the heavier prong leads to a (small) wrong-branch contribution. This can be avoided by discarding the subjet candidate with smaller transverse mass $m_\perp^2 \equiv m^2 + p_\perp^2$ instead. This modification is known as modified mass-drop tagger (mMDT).

ATLAS default top tagger [86], [87]: Building on k_T splitting scales and the `YSplitter`, the `ATLAS default top tagger` reveals further information on the clustering history. The first two scales y_{1-2} and y_{2-3} correspond to the subsequent decays $t \to W^+b$ and $W \to jj$. A multivariate cut on y_{1-2}, y_{2-3}, y_{3-4} as well as jet mass and transverese momentum gives good discrimination against QCD jets. In recent implementations, all splitting scales are processed through a neuronal network.

Johns Hopkins tagger [89] *and CMS top tagger* [88], [129]: The `Johns Hopkins tagger` uses a C/A fat jet of radius $R = 0.8$ as input, giving good efficiencies for top transverse momenta around $p_\perp \sim 800\,\text{GeV}$. It derives from the `BDRS Higgs tagger`, implementing several important changes due to the two-step decay of the top quark. The algorithm does not stop after the first splitting has been identified, but both subjets are kept and are candidates for an additional splitting. This results in three or four subjets for successful candidates. Instead of mass drop and a symmetry criterion, the requirement is formulated in terms of transverse momentum ($\min[p_{i\perp}, p_{j\perp}]/p_{\perp,\text{fatjet}} > \delta_p$) and angular separation ($|\Delta\eta_{ij}| + |\Delta\phi_{ij}| < \delta_r$). Cuts are applied on the top candidate mass $m_{jjj(j)}$, the mass of the best W candidate m_{jj}, and the helicity angle θ_h. The latter is the opening angle between the top momentum and the softer of the two W decay subjets, measured in the rest frame of the W boson candidate.

A modified version of the algorithm is known as the `CMS top tagger`. It uses a slightly different splitting criterion based on R separation. In addition, the cuts on the W candidate mass and θ_h are replaced by requiring all pairwise masses of the three leading subjets to fulfill $m_{jj} > 50\,\text{GeV}$.

HEPTopTagger [24]: The `HEPTopTagger` (where "HEP" stands for "Heidelberg–Eugene–Paris") also generalizes the `BDRS Higgs tagger`. Because of its relevance to the original parts of this thesis, we describe the algorithm in more detail for later reference. The `HEPTopTagger` starts from C/A fat jets with a large radius $R = 1.5$ to access top quarks down to very moderate boosts $p_\perp \gtrsim 200\,\text{GeV}$. The following procedure is imposed on the fat jet:

1. Undo the last clustering of the jet j into j_1, j_2, ordered $m_{j_1} > m_{j_2}$.

2. If a significant mass drop occurred, $m_{j_1} < \theta \cdot m_j$, both j_1 and j_2 are kept as candidate subjets. Otherwise discard j_2.[6]

3. Repeat these steps for the kept subjets unless $m_{j_i} < \mu$, in which case j_i is added to the set of output subjets.

4. For all three-subjet combinations, apply a filtering stage (cf. Sec. 3.3) using $R_\text{filt} = \min(\,0.3, \min \Delta R_{ij}\,)$. The five hardest subjets are then reclustered to exactly three filtered subjets. Choose the combination whose filtered mass is closest to the true top mass m_t as top candidate. Tagging fails if less than three subjets are found or the candidate mass is not within $m_t \pm 25\,\text{GeV}$.

5. The three subjets are ordered by transverse momentum, $p_{\perp 1} > p_{\perp 2} > p_{\perp 3}$. Their pairwise masses (m_{12}, m_{13}, m_{23}) have to satisfy one of the following three conditions known as the A-cut (because

[6] Note that one could again define a modified mass-drop tagger similar to footnote 5. As this modification is irrelevant for the main parts of this thesis, we do not further follow this notion.

they form the letter "A" if m_{23}/m_{123} is plotted against $\arctan m_{13}/m_{12}$):

$$\begin{aligned}
&\left[0.2 < \arctan\frac{m_{13}}{m_{12}} < 1.3 \qquad \text{and} \qquad R_{\min} < \frac{m_{23}}{m_{123}} < R_{\max}\right] \\
\text{or} \quad &\left[R_{\min}^2\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\max}^2\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right)\right. \\
&\qquad \left. \text{and} \qquad \frac{m_{23}}{m_{123}} > R_{\text{soft}}\right] \\
\text{or} \quad &\left[R_{\min}^2\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\max}^2\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right)\right. \\
&\qquad \left. \text{and} \qquad \frac{m_{23}}{m_{123}} > R_{\text{soft}}\right] .
\end{aligned} \tag{3.10}$$

Default cut parameters are $R_{\min} = 0.85\,\frac{m_W}{m_t}$, $R_{\max} = 1.15\,\frac{m_W}{m_t}$, and $R_{\text{soft}} = 0.35$.

6. For consistency, the top candidate must satisfy $p_\perp \geq 200\,\text{GeV}$.

The mass-drop (MD) procedure (steps 1–3) serves two purposes: It grooms the jet from (large-angle) soft radiation and applies a criterion to identify a non-specified number of separate prongs based on jet mass. The default values of the two free parameters are chosen as $\theta = 0.8$ and $\mu = 30\,\text{GeV}$ [24], [130]. The filtering stage makes the `HEPTopTagger` (and all other `BDRS` derivatives) robust against additional soft radiation from pile-up or underlying event as discussed in Sec. 3.3. Due to the mass-drop procedure with no reference to an external mass or energy scale nor a fixed number of subjets, the algorithm is also robust against splash-in from other hard jets. As long as the top subjets are identified correctly all other subjets will be ignored.

We only briefly described a small and subjective selection out of the bountiful mass of recent tagging methods. For a broader overview and more details the reader is referred to the existing excellent reviews on the topic [18], [20], [21], [131].

Note that we only focused on those tagging algorithms for boosted top quarks that rely on the notion of a large-radius fat jet and its substructure analysis. Indeed this class constitutes the vast majority of recent top taggers[7]. These taggers often involve a grooming stage to reduce the effects of soft QCD described in Sec. 3.2. They are also naturally vulnerable to splash-in from unrelated hard partons, hence they mainly populate the low-multiplicity region in Fig. 3.2.

[7] Notable exceptions are shower / event deconstruction [132], [133] and template overlap [134].

— I could offer a million answers - all false.

Mark Renton

4

More flexibility with jet clustering with a terminating veto

This chapter is based on work by the author [A, E].

We introduce a new class of jet clustering algorithms. In these algorithms, a criterion inspired by successful mass-drop taggers is applied that prevents the recombination of two hard prongs if their combined jet mass is substantially larger than the masses of the separate prongs. This "mass jump" veto effectively results in jets with variable radii in dense environments. We investigate differences to existing methods. For boosted top quarks we show that the new algorithm has beneficial properties, which can lead to improved tagging purity.

4.1 Introduction

Many different jet algorithms have been proposed so far. In Sec. 3.2 we reviewed the most commonly used ones and commented on the problem of choosing the optimal radius parameter R. As the majority of algorithms construct jets with fixed R, this choice is always a compromise that has to balance between splash-in and splash-out. The authors of Ref. [135] point out that jet clustering with a radius depending on the jet momentum can be advantageous and give more accurate results. They consequently propose to cluster with $R \propto 1/p_\perp$, resulting in jets with variable radii. An entirely different approach to the same problem is taken by mass-drop tagging algorithms like the `BDRS Higgs tagger` or the `HEPTopTagger` (cf. Secs. 3.5 and 3.6). They address heavy resonances that are so highly boosted that their subsequent

decay products cannot reasonably be resolved with conventional jet algorithms. Due to the high centre-of-mass energy of the LHC, boosted top quarks, Higgs bosons, etc. are expected to be produced in larger numbers during the upcoming run. To identify these resonances, it is possible to capture all decay products in a large-radius fat jet and apply substructure methods. In short, the basic idea states that a jet should be broken up into two separate subjets if the jet mass experiences a significant drop in the procedure. These algorithms identify hard substructure without referring to a fixed (sub)jet radius and turned out to perform very well in Higgs boson and top quark tagging (see e.g. Refs. [18], [19], [131] for reviews). Implicitly, a $p_\perp$-dependent subjet radius is given by the mass cut, as the characteristic separation between the daughters of an energetic resonance is $\Delta R_{\text{daughters}} \gtrsim 2m_{\text{mother}}/p_\perp$.

In this chapter, we supplement existing jet algorithms with a recombination veto, which may prevent further clustering at a jet radius smaller than the given R. The working principle is similar to mass-drop tagging: if the recombination of two protojets leads to a significant mass jump, they should be resolved separately. In contrast to other algorithms with variable radii, the veto is a property of two jets, i.e. the effective clustering radius now also depends on the jet's vicinity. In this way well-separated jets are clustered conventionally with only small deviations, whereas on the other hand the merging of two hard prongs into a heavy resonance is vetoed.

The introduction of a clustering veto is not a novelty. One notable example is given by pruning (cf. Sec. 3.3), which follows a different philosophy, though. There a recombination step is vetoed if it resembles large-angle soft radiation (expressed in terms of transverse momentum and R separation) in the sense that jet clustering proceeds as usual after the softer part has been discarded from the event. Using this procedure only hard substructure is kept and the algorithm can already be used as a tagger. In contrast, here we suggest what we call a terminating veto for the mass-jump procedure: when the merging of two hard prongs is vetoed, they no longer participate in jet clustering. By this procedure (sub)jets are identified without reference to an external energy or angular scale, while keeping all the radiation present in the event.

4.2 The algorithm

4.2.1 A comment on mass-drop unclustering

We reviewed the mass-drop (MD) unclustering procedure in the context of the `HEPTopTagger` in Sec. 3.6. The crucial criterion here is the mass drop: if a splitting $j \to j_1, j_2$ satisfies $m_{j_1} < \theta \cdot m_j$, both subjets are kept as hard substructure; otherwise the prong with a smaller invariant mass j_2 is discarded. Subjets are not unclustered any further if their jet mass is below the scale μ.

Note that the un-clustering algorithm is designed to follow the cascade decay chain of the top quark,

$$t \to bW^+ \to bq\bar{q}' \,. \tag{4.1}$$

At parton level the successive mass drops $\tau = \frac{m_{j_1}}{m_j}$ are given by

$$\tau_1 = \frac{m_W}{m_t} \approx 0.46 \,, \quad \tau_2 = \frac{m_q}{m_W} \approx 0 \,, \tag{4.2}$$

hence the parameter θ has to be chosen sufficiently large to incorporate the first decay (the default value is $\theta = 0.8$ [24], [130]). This is in contrast to the mass-jump clustering algorithm to be introduced below. There, the equivalent of the first mass drop is absent and all mass ratios τ are expected to be of same order. In Sec. 4.3.2 we argue that this feature is beneficial for boosted top quark tagging.

We note that in case the unclustering proceeds via $t \to j(bj) \to jbj$ one obtains

$$\tau_1' = \frac{\sqrt{m_t^2 - m_W^2}}{2m_t} \Delta R_{bj} \,, \tag{4.3}$$

which is typically smaller than τ_1. $\Delta R_{bj} = \sqrt{\Delta y^2 + \Delta\phi^2}$ is the R-distance between the subjets b and j.

4.2.2 The mass-jump clustering algorithm

In the following we present the mass-jump (MJ) clustering algorithm. This algorithm is a modification of the infrared and collinearly safe generalized-k_T clustering algorithms reviewed in Sec. 3.2. They define a distance measure d,

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min\left[p_{i\perp}^{2n}, p_{j\perp}^{2n}\right] \,, \quad d_{iB} = p_{i\perp}^{2n} \,. \tag{3.1}$$

In the commonly used generalized-k_T algorithms, it serves two purposes. First, it is an ordering variable: it determines the order of recombination given by the pair with the smallest distance d_{ij} at each step. Second, it is a test variable: it acts as an upper bound on the jet radius, because a minimal beam distance d_{iB} implies $\Delta R_{in} > R \;\forall$ protojets n.

To construct the mass-jump clustering algorithm, we modify the test variable by introducing an additional veto.[1] In the spirit of a reverse mass-*drop* procedure, "sub"jets are directly constructed by examining the veto condition at each recombination step, where the parameter θ now acts as a mass-*jump* threshold. First, all input particles are labelled as *active* protojets. The recombination algorithm is defined as follows:

1. Find the smallest d_{ij} among active protojets; if it is given by a beam distance, d_{iB}, label i *passive* and repeat step 1.

2. Combine $ij \to k$ by summing their four-momenta, $p_k = p_i + p_j$ (E-scheme). If the new jet is still light, $m_k < \mu$, replace i and j by k in the set of active protojets and go back to step 1.
   Otherwise check the mass-jump criterion: if $\theta \cdot m_k > \max\left[m_i, m_j\right]$ label i and j *passive* and go back to step 1.

3. Mass jumps can also appear between an active and a passive protojet. To examine this

   a. Find the passive protojet n that is closest to i in terms of the metric d and is not isolated, $d_{in} < d_{nB}$.

   b. Then check if these two protojets would have been recombined if n had not been rendered passive by a previous veto, i.e. $d_{in} < d_{ij}$.

[1] Separate measures for ordering variable and test (veto) variable were first introduced in Ref. [101].

c. Finally check the mass-jump criterion, $m_{i+n} \geq \mu$ and $\theta \cdot m_{i+n} > \max[m_i, m_n]$.

If all these criteria for the veto are fulfilled, label i *passive*. Do the same for j. If either of i or j turned passive, go back to step 1.

4. No mass jump has been found, so replace i and j by k in the set of active protojets. Go back to step 1.

Clustering terminates when there are no more active protojets left. Passive protojets are then labelled jets. Note that for $\theta = 0$ or $\mu = \infty$ standard sequential clustering without veto is recovered, which in this case can be reduced to steps 1 and 4.

The mass-jump algorithm described above is the first member of the family of jet clustering algorithms with a terminating veto and has been made publicly available as part of the FastJet contribution package [E]. The plugin is dubbed `ClusteringVetoPlugin` and accepts any user-defined veto function. Its exemplary usage is illustrated within the package. and in Appendix A.1.

4.2.3 Properties

The mass-jump veto only has an impact on protojets that are separated by $\Delta R < R$ and whose combined mass would be above the (arbitrary) scale μ. It is designed to resolve close-by jets (which could come from the decay of a boosted resonance such as t, $W^{\pm}$, Z, H, ...) separately. As the vetoed jets are excluded from further clustering, their effective jet radius is smaller than the parameter R, which now gives an upper bound. A lower bound is indirectly induced by a finite threshold scale μ.

There are several similarities and differences compared to MD unclustering. Fig. 4.1 schematically depicts a standard clustering sequence (e.g. of a hadronically decaying boosted top quark) and how the two algorithms act on the given event. The clustering sequence is to be read from right to left; hard prongs are depicted as straight lines, whereas wiggly lines symbolize soft radiation. The mass-drop tagger (MDT) sequentially unclusters a fat jet (which can be an actual large-radius jet or the whole event) from left to right, whereas the MJ algorithm starts from the fat jet's constituents and proceeds to the left. The final (sub)jets are indicated by red cones.

In the MDT algorithm (upper panel), starting from a fat jet soft radiation is groomed away (1) until at one unclustering step the mass-drop criterion is fulfilled, resulting in two subjets (2). The same grooming–tagging procedure continues for every prong that experiences a further mass drop (3+4). More soft radiation is removed (5) until the subjet masses are below the threshold μ (6). The remaining prongs are now labelled "subjets".

MJ clustering (lower panel), on the other hand, is identical to standard clustering algorithms until the jet mass exceeds μ (a).[2] Clustering continues (b) until the next recombination step would result in a substantial mass jump (c), at which step clustering is vetoed and the two prongs turn passive. Active protojets continue clustering (d) unless a veto is called, which can also act against a (hypothetical) recombination with a passive jet (e). Jet clustering continues for the remaining particles, giving additional jets (f).

[2] Or the jet has reached its size given by the radius R – for the sake of comparison with the MDT procedure, we take $R = \infty$ for the moment.

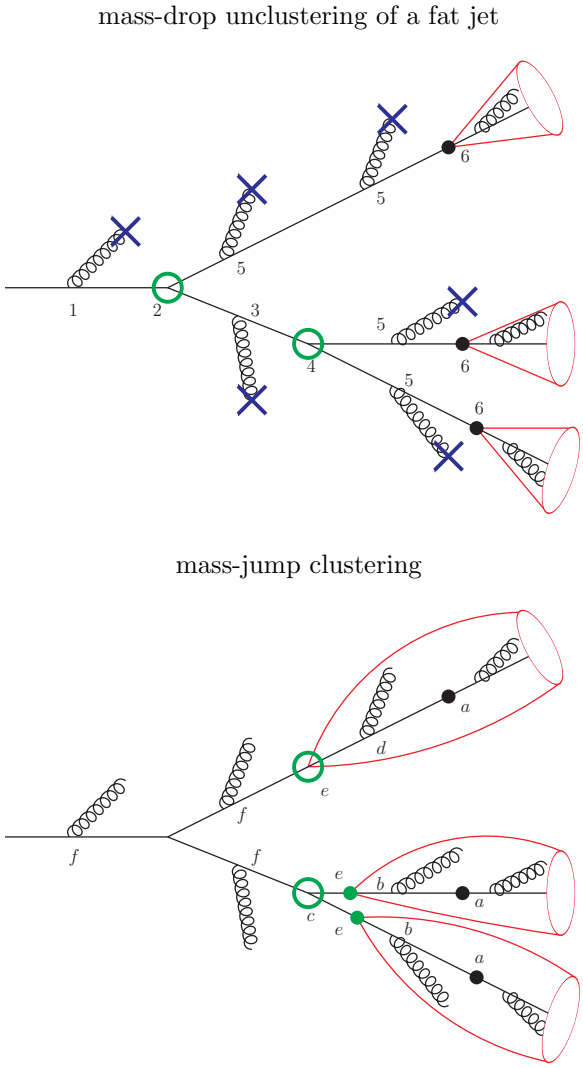


Figure 4.1: Key differences between MD unclustering (top) and MJ clustering (bottom) are visualized for a schematic clustering sequence (e.g. of a hadronically decaying boosted top quark). Sequential recombination is performed starting from the constituents at the right-hand side, such that in the upper panel the very left line symbolizes the whole fat jet, which is then sequentially unclustered again (bottom panel: MJ clustering works its way from the constituent particles to the left). Inside the cluster sequence, hard prongs are depicted as straight lines, whereas wiggly lines symbolize soft radiation. Black dots denote the jet mass threshold $m = \mu$, and green circles indicate a mass drop (or mass jump). The final (sub)jets are indicated with red cones. The individual steps of the respective two algorithms (steps 1–6 for MD unclustering, steps a–f for MJ clustering) are described in the text.

In the idealized case, the output jets of both algorithms are comparable but differ in two aspects. First, MDT subjets are groomed even after a mass drop until they reach $m < \mu$ whereas MJ jets continue collecting radiation in the regime between $m > \mu$ and the mass jump. Although this effect is expected to be absent for reasonably large values of μ, if undesired it is straightforward to apply MDT-like grooming on the MJ jets. Second, the MJ clustering algorithm also returns jets that did not experience mass jumps (f) that are absent among MDT subjets (1,3,5). These can be desirable (well-separated jets for finite R) or can be considered junk; in the latter case it is again straightforward to remove them as these are the only jets turned *passive* by the upper bound on the jet radius instead of a mass jump.

Also note the important property that MD unclustering experiences cascade mass drops (cf. Sec. 4.2.1) while MJ clustering does not. This results in all mass jumps being among single hard prongs with a typical scale $\sim m_{\text{heavy resonance}}/\mu$, i.e. the threshold parameter θ can be chosen substantially lower.

4.3 Performance

4.3.1 Sparse environment: QCD dijets

We compare the MJ clustering algorithm to its standard counterparts. QCD dijet events are expected to contain two well-separated hard jets, but more jets may be found due to large-angle emissions or jet substructure induced by the parton shower. In particular MJ clustering is prone to misidentify jet substructure as separate hard objects, and this section aims to quantify this effect of the veto. 10,000 QCD dijet events are simulated with Pythia 8.186 [136] where the minimum parton transverse momentum at matrix element level is chosen $\hat{p}_\perp^{\min} = 40\,\text{GeV}$. The analysis is implemented as a Rivet 2.1.2 [137] plugin.

Jets are constructed from all (visible) final-state particles with pseudo-rapidity $|\eta| \leq 4.9$. The clustering parameters are chosen $R = 0.8$ and $p_\perp \geq p_\perp^{\min} = 50\,\text{GeV}$, also jets are required to be sufficiently central, $|\eta| \leq 4.0$. We compare the jets clustered with a standard algorithm (anti-k_T, Cambridge/Aachen, or k_T algorithms as provided by FastJet 3.0.6 [138], [139]) to those obtained with the corresponding MJ algorithm on an event-by-event basis. Only events that contain at least one hard jet from the standard algorithm, $p_\perp^{\text{std}(1)} \geq 150\,\text{GeV}$, are accepted. This assures that the leading jet is still present among the MJ jets and does not drop below $p_\perp^{\min}$, even if torn apart by the clustering veto. For each of the algorithms, ~ 100 events pass this cut.

The three standard algorithms agree very well in the number of jets n_{std}, which is 2 (in roughly one in two events) or above. We perform a parameter scan for the MJ clustering arguments θ and μ. Fig. 4.2 (bottom panel) shows the difference in the average number of jets per event ($\Delta\bar{n} = \bar{n}_{\text{MJ}} - \bar{n}_{\text{std}}$). The mutual leading jets (i.e. the $\min[n_{\text{MJ}}, n_{\text{std}}]$ jets with largest $p_\perp$) in each event are matched, and differences between the MJ and standard algorithms are investigated on a jet-by-jet basis. For each pair $(j_{\text{MJ}}, j_{\text{std}})$, we obtain the R-distance ($\Delta R_{j_{\text{MJ}}, j_{\text{std}}}$) and relative difference in transverse momentum ($\Delta p_\perp / p_\perp^{\text{std}} = \frac{p_\perp^{\text{std}} - p_\perp^{\text{MJ}}}{p_\perp^{\text{std}}}$). The upper two panels of Fig. 4.2 show the values of these two observables averaged over all matched jet pairs. For large parts of the parameter space, the effects of the clustering veto are only limited in the QCD dijet scenario considered here.

Differences between individual jets (upper two rows) are negligibly small in the small-θ and large-μ parameter regions for all three jet algorithms. This behaviour is expected as these are the limits where

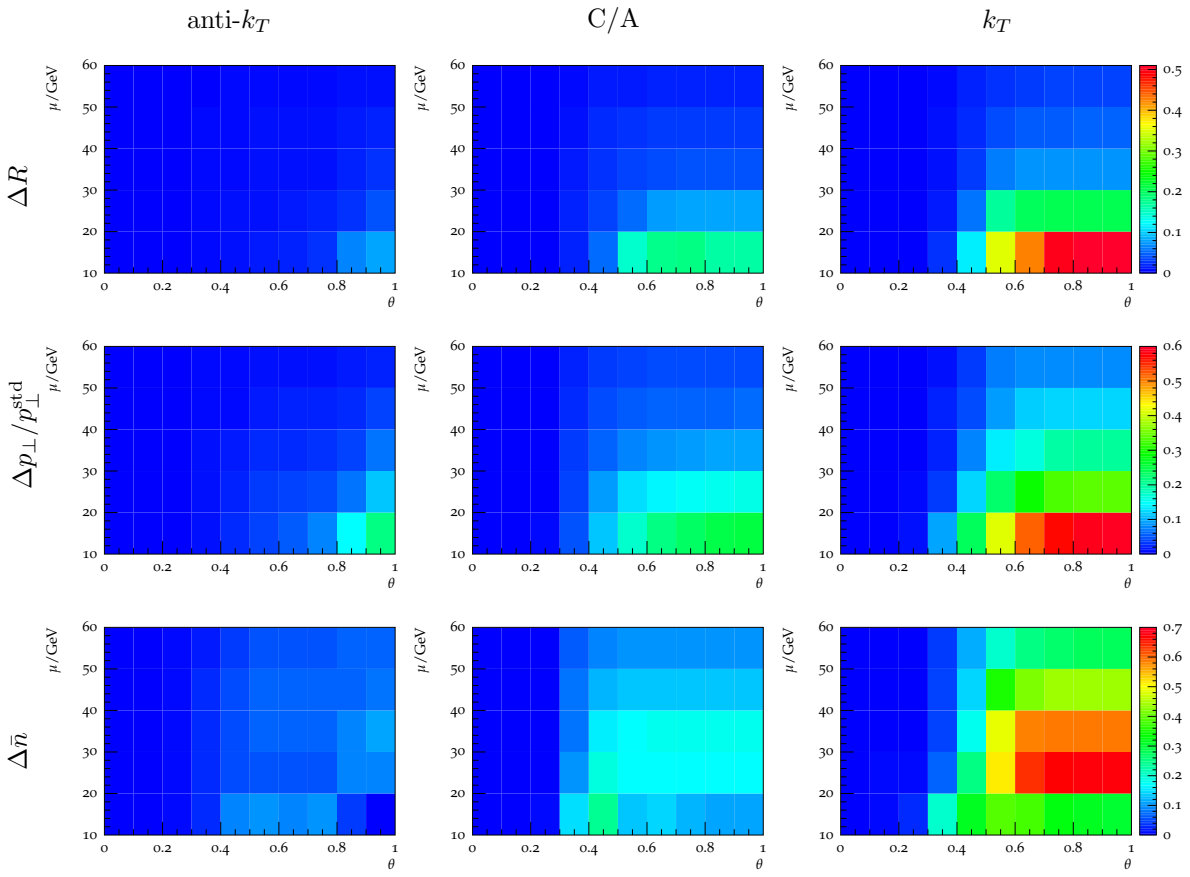


Figure 4.2: Comparison between MJ clustering and its standard counterparts for the anti-k_T (left), C/A (middle) and k_T (right) algorithms. All jets were clustered with $R = 0.8$ and $p_\perp \geq 50\,\text{GeV}$, and only events where $p_\perp^{\text{std}(1)} \geq 150\,\text{GeV}$ were accepted. The averaged values of the three observables ΔR, $\Delta p_\perp / p_\perp^{\text{std}}$ and $\Delta \bar{n}$ are shown from top to bottom for a range of parameters θ and μ.

the veto is rendered ineffective. The closer the parameters are chosen to the strong-veto region ($\theta \to 1$, $\mu \to 0$), deviations between the vetoed and standard algorithms grow larger. In particular for the k_T algorithm these differences can be substantial, namely $\Delta R \sim 0.5$ and $\Delta p_\perp / p_\perp^{\text{std}} \sim 0.6$ for the considered setup. The C/A and especially the anti-k_T algorithm behave much more moderately under the MJ veto. For the latter, deviations only reach $\Delta R \sim 0.1$ and $\Delta p_\perp / p_\perp^{\text{std}} \sim 0.2$ even in the strong-veto region, and are almost absent in the bulk of parameter space.

Generally the differences between MJ-vetoed and standard clustering are smallest for the anti-k_T algorithm and largest for the k_T algorithm, with the C/A algorithm taking an intermediate position. This characteristic is directly related to the ordering of the cluster sequence (cf. Sec. 3.5), which is crucial in the MJ algorithm. If soft particles are clustered first (k_T), it is very likely to induce fake substructure that will fulfill the mass-jump condition at the stage when these soft clusters are recombined. The anti-k_T algorithm on the other hand ignores the parton showering history and clusters around hard prongs. It is therefore much more robust, while the purely angular-based C/A algorithm is moderately prone to vetoing fake soft clusters.

The number of jets is naturally equal or larger in the vetoed algorithms compared to the standard algorithms with equal jet clustering radius (Fig. 4.2 lower panels). If, however, the veto acts too strong, hard jets are split and may not pass the $p_\perp \geq p_\perp^{\text{min}}$ cut any more, resulting in a decreasing number of jets again. For large minimum jet transverse momentum close to $p_\perp^{\text{std}(1)}$, say $p_\perp^{\text{min}} = 100\,\text{GeV}$ for our analysis, $\Delta\bar{n}$ can ultimately become negative.

Also for other jet clustering radii and $p_\perp$ thresholds, results are qualitatively very similar to the ones described above. We therefore omit further plots.

4.3.2 Busy environment: boosted top quarks

Tagging boosted top quarks is an important target in many current experimental studies and also an ideal playground to investigate the performance of MJ clustering in busy environments. In order to probe the moderately boosted energy regime and illustrate the algorithm, we simulate top pair production via a hypothetical heavy vector boson,

$$pp \to Z' \to t\bar{t} \to \text{hadrons} \tag{4.4}$$

for three different resonance masses $m_{Z'} = 500\,\text{GeV}$, 700 GeV, and 1 TeV. Using the Cambridge/Aachen algorithm we construct fat jets with radius $R = 1.5$ and transverse momentum $p_\perp \geq 200\,\text{GeV}$. The first sample with a relatively light Z' results in fat jets whose $p_\perp$ distribution drops steeply to mimic top quarks produced in SM processes. The latter two samples emulate a generic heavy resonance and yield top quarks with transverse momenta peaking around ~ 300 and $\sim 450\,\text{GeV}$, respectively. The fat jets are fed to the `HEPTopTagger`, which performs the following three-step procedure (for details see Sec. 3.6).

i. Subjets are obtained from the fat jet via mass-drop unclustering.

ii. The subjets are filtered (cf. Sec. 3.3).

iii. Cuts on subjet mass ratios and the candidate's transverse momentum are applied.

For comparison with our veto algorithm, we apply the same `HEPTopTagger` algorithm but where the subjets are now obtained directly with MJ clustering, starting from the fat jet's constituent particles. Steps ii and iii remain unchanged such that the difference in tagging performance can be directly compared. We take $R = \infty$ and scan the parameter space in θ and μ. Results are based on each 10,000 signal and background events (QCD dijets with $\hat{p}_\perp^{\min} = 150\,\text{GeV}$) generated with PYTHIA 8 and analyzed within RIVET. The resulting tagging efficiencies $\epsilon = \frac{\#\text{tags}}{\#\text{fat jets}}$ are shown in Fig. 4.3.[3]

Indeed the peak tagging efficiencies are equal for both algorithms and constant over a relatively large part of parameter space. However, as argued in Sec. 4.2.3, mass-jump jet finding allows for well-performing top tagging in a much wider range in the parameter θ. The reason for this behaviour lies in the absence of an equivalent to the cascade mass drops experienced in MDT's (such as $t \to bW^+ \to bjj$). This feature can also be directly seen in Fig. 4.3 where in the MDT case (left) the onset of top tagging is around $\theta = 0.5 \approx \frac{m_W}{m_t}$, whereas for MJ clustering (right) the characteristic scale is much lower. In particular, lower values of θ correspond to a much stricter identification of separate jets, which might turn out beneficial for background rejection.

A comparison of the plots reveals some additional interesting features. For MD unclustering (left), a low parameter θ increases the possibility of too strong grooming. This effect is further enhanced for low μ, which leads to a rounded corner in the available parameter space. In the MJ algorithm on the other hand, non-vetoed prongs are merged instead of groomed for low θ. Very low values of μ allow additional (unphysical) vetoes among soft prongs. Both effects separately worsen the correspondance between the subjets and the MC truth quarks. As they are complimentary in terms of subjet multiplicity, however, their delicate interplay leads to the complicated structure of the available parameter space.

A feature common to both algorithms is the decreased dependence on μ for higher top quark boosts. Jet mass is a Lorentz invariant and theoretically expected to peak around $m^2 \sim p_\perp^2 R^2\, e^{-1/\sqrt{\alpha_S C_F/\pi}}$ for isolated quark jets [119].[4] With a top boost around 400 GeV, a rough estimate for the softest (leading) subjet with $p_\perp \simeq 50\,\text{GeV}$ (250 GeV) and radius $R \simeq 0.6$ (see also Fig. 6.1 for the correlation between top boost and radial distances) yields $m \simeq 3\,\text{GeV}$ (16 GeV) where we used $\alpha_S = 0.12$ and $C_F = 4/3$. Another source of jet mass is uncorrelated radiation from nearby partons or the underlying event. Eq. (3.2) estimates the latter to contribute $\delta m \simeq 8\,\text{GeV}$. The lower plots of Fig. 4.3 suggest that μ should be chosen larger than the mass of the softest subjet, which is dominated by soft uncorrelated radiation. For mild boosts, the characteristic subjet radius increases as $1/p_\perp$ and for $p_\perp \simeq 30\,\text{GeV}$ (130 GeV) and $R \simeq 1$ we obtain estimates $m \simeq 3\,\text{GeV}$ (14 GeV) and $\delta m \simeq 17\,\text{GeV}$ respectively, consistent with the low-μ region in the upper plots.

The effect of jet contamination cancels to some extent in mass ratios, which makes the dependence on θ independent of the top quark boost. From the results obtained with MJ clustering (right) we can get another estimate of the scale of subjet masses. The onset value $\theta \sim 0.3$ yields $m \sim 0.3\, m_W \sim 25 - 30\,\text{GeV}$ depending on the contribution from uncorrelated radiation to the W jet mass, consistent with our estimates for the leading subjet above.

[3] Fat jets that deviate too much from their Monte Carlo truth top quark ($\Delta R_{j^{\text{fat}},t^{\text{MC}}} > 0.6$) are ignored in signal events.

[4] Note that the characteristic angular distances are inversely proportional to the relevant object's transverse momentum. Hence boost invariance of jet mass is manifest in this expression if the jet radius is adjusted accordingly.

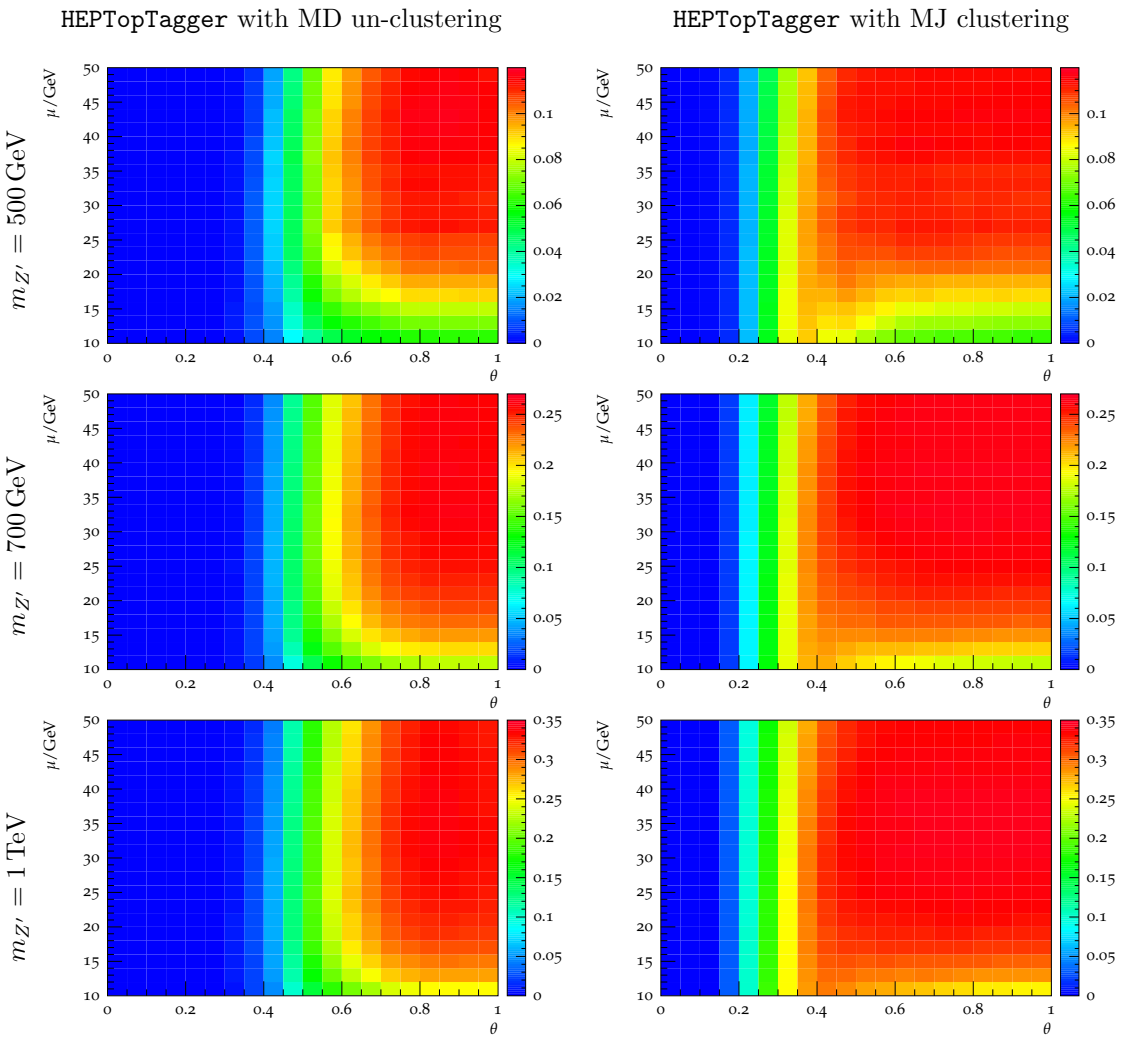


Figure 4.3: Top tagging efficiency ϵ for the `HEPTopTagger` with MD un-clustering (left) and MJ clustering (right). For both algorithms the parameter space θ, μ is scanned. From top to bottom, the panels show signal rates for the $m_{Z'} = 500\,\text{GeV}$, 700 GeV, and 1 TeV samples.

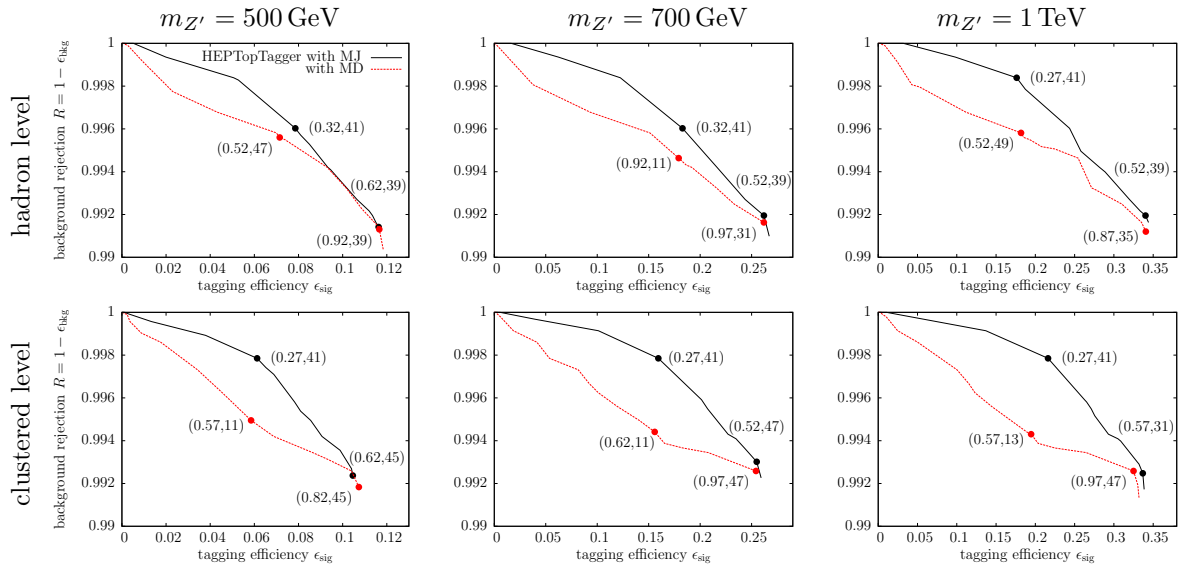


Figure 4.4: Receiver-operating characteristic (ROC) curves for top tagging using the `HEPTopTagger`. Subjet finding with the MJ clustering algorithm (black solid) is compared to the original algorithm, which employs MD un-clustering (red dashed). From left to right, the upper panels show results at hadron level for the $m_{Z'} = 500\,\text{GeV}$, $700\,\text{GeV}$ and $1\,\text{TeV}$ sample, respectively. The lower panels are similar but obtained from hadrons centred into cells of size 0.1×0.1 in η–ϕ space to emulate finite detector resolution. Parameters at exemplary benchmark points are given for illustration in the format $(\theta, \mu/\text{GeV})$. If high purity is desired, MJ clustering gives improved performance.

The observed overall increase in tagging efficiency for larger resonance masses $m_{Z'}$ is a simple consequence of the underlying kinematics. The majority of fat jets carry a larger transverse momentum than the respective initiating top quark. As a result, the very last cut ($p_\perp^{\text{top candidate}} \geq 200\,\text{GeV}$) rejects many moderately-boosted candidates even in the case of perfect reconstruction. With larger boost (corresponding to larger $m_{Z'}$), this fraction becomes smaller.

Fig. 4.4 compares the receiver-operating characteristic (ROC) curves of the original `HEPTopTagger` and the modified algorithm where MD unclustering has been replaced by MJ clustering.[5] It is observed that signal tagging efficiency and background rejection coincide for large efficiencies, giving $\epsilon_{\text{sig}} \approx 0.12$ and $R = 1 - \epsilon_{\text{bkg}} \approx 0.991$ for the $m_{Z'} = 500\,\text{GeV}$ sample, $(0.26, 0.991)$ for the $m_{Z'} = 700\,\text{GeV}$ sample, and $(0.34, 0.992)$ for the $m_{Z'} = 1\,\text{TeV}$ sample, respectively. These values correspond to the plateau at large θ and medium-to-large μ in Fig. 4.3. However, due to the enlarged parameter space, the MJ algorithm outperforms the standard procedure and should be preferred in the transition (high-purity) region. This result is even more pronounced if limited detector resolution is taken into account. For our

[5] These curves are obtained from the full parameter scan. Among all setups (θ, μ) that give a similar signal tagging efficiency, only the one that yields the highest background rejection is picked and plotted. Note that different setups can yield very similar efficiencies, and that the benchmark points given in the figure are chosen somewhat arbitrarily in this sense.

simple analysis, this is implemented by applying a cellular grid in the η–ϕ plane (cell size 0.1×0.1) and replacing all stable hadrons to the centre of their respective cells. For most working points, the inevitable decrease in performance is less pronounced when MJ clustering is used. At maximum tagging efficiencies, the two algorithms still give the same results.

4.4 Conclusions

We developed and investigated a new jet clustering algorithm that includes a recombination veto based on jet mass. In this mass-jump (MJ) procedure, the clustering radius R now acts as an upper limit on jet size and the merging of two hard prongs is prevented. We showed that in sparse events with well-separated jets, the effects of the veto are very limited in a large range of the parameter space. Also the anti-k_T clustering algorithm is more robust against fake two-prong substructure than the Cambridge/Aachen and k_T algorithms. In the dense environment of hadronically decaying boosted top quarks, MJ clustering gives results comparable to those of mass-drop taggers (MDT) by which the veto was inspired in the first place; the main difference being that cascade mass drops as present in MDT's are avoided, which in turn allows for stricter threshold parameters. The larger parameter space then leads to improved ROC curves for the `HEPTopTagger` when the mass-drop procedure is replaced by MJ clustering.

Until the veto is interposed, MJ jet clustering proceeds identically to its standard counterpart. In particular, no soft radiation is removed and after the veto (multiple vetoes) additional jets are formed from the remaining particles. Especially in realistic scenarios when soft QCD radiation (from underlying event or pile-up) is present, the application of grooming techniques (cf. Sec. 3.3) can improve jet shape observables by removing these uncorrelated energy deposits.

Jet algorithms with a terminating veto are a promising tool for collider experiments as they make room for more flexibility. The optimal clustering radius depends on various parameters such as the type of initiating particle, its energy or transverse momentum, and the surrounding topology of the event. The MJ veto automatically adjusts the jet radius such that hard substructure is separated into isolated jets. This feature may prove helpful in a variety of events where jets are not well-separated and is investigated in a realistic process in the next chapter.

— If you do not dance you have no purpose.

Vera Gorski

5

Boosted top quarks in a very busy environment

This chapter is based on work by the author [B].

At the LHC, tagging boosted heavy particle resonances which decay hadronically, such as top quarks and Higgs bosons, can play an essential role in new physics searches. In events with high multiplicity, however, the standard approach to tag boosted resonances by a large-radius fat jet becomes difficult because the resonances are not well separated from other hard radiation. In this chapter, we propose a different approach to tag and reconstruct boosted resonances by using the mass-jump jet algorithm of Chapter 4. A key feature of the algorithm is the flexible radius of the jets, which results from a terminating veto that prevents the recombination of two hard prongs if their combined jet mass is substantially larger than the masses of the separate prongs. The idea of collecting jets in "buckets" is also used. As an example, we consider the fully hadronic final state of pair-produced vectorlike top partners at the LHC, $pp \to T\bar{T} \to t\bar{t}HH$, and show that the new approach works well. We also show that tagging and kinematic reconstruction of boosted top quarks and Higgs bosons are possible with good quality even in these very busy final states. The vectorlike top partners are kinematically reconstructed, which allows their direct mass measurement.

5.1 Introduction

The prospect of boosted heavy resonances at the LHC has sparked the whole field of tagging algorithms based on fat jet substructure described in Secs. 3.5 and 3.6. An even more challenging situation that is accessible in the upcoming run arises when multiple heavy resonances are produced simultaneously.

Such processes lead to very busy final states where the heavy particles under consideration (such as top and Higgs) as well as their daughter particles are not well separated. As a result, fat jets merge in most events, and a majority of the jets contain decay products from more than one resonance. Such a scenario is not adequately addressed by most tagging algorithms based on (isolated) fat jets.

In this chapter, we suggest a new framework of jet tagging that allows particle reconstruction with good quality compared to traditional methods in such a busy hadronic environment. A key ingredient is the mass-jump jet clustering algorithm developed in Chapter 4. There it has been shown that the algorithm gives competitive performance when employed in the `HEPTopTagger`, but now the "sub"jets are formed directly without the definition of an intermediate fat jet. Mass-jump clustering harnesses the advantages of fat jet substructure algorithms of resolving small jets without reference to a fixed radius.

The absence of the fat jet, however, reintroduces the issue of large combinatorics in such busy environments, since the decay products of the heavy resonances cannot be disentangled a priori. To facilitate event reconstruction, the idea of collecting jets into separate "buckets" [140], [141] is applied, which allows efficient assignment of jets to their respective resonances.

The jet-tagging method proposed here is applicable to a broad range of Standard Model and BSM phenomena at hadron colliders. There are indeed important Standard Model processes which involve decays of multiple heavy particles, resulting in a busy hadronic environment. A prime example is the associated production of a Higgs boson with two top quarks ($pp \to t\bar{t}H$), which has attracted attention as this channel opens up the opportunity to measure directly the Higgs–top Yukawa coupling, an essential probe toward understanding the Higgs sector.

Some models of supersymmetry also predict large multiplicity of jets with little or no missing transverse energy (MET). For example, assuming that the gluino is the lightest SUSY particle, it can decay into a top quark and jets when baryonic R-parity associated to the third generation quark is violated [142]. This leads to a multijet final state when the top quark decays hadronically ($\tilde{g}\tilde{g} \to ttjjjj$). Another example is the stealth SUSY, where the top and the lighter stop ($\tilde{t}_1$) masses are almost degenerate, leading to final states without significant MET [143]. The heavier stop ($\tilde{t}_2$) has a model-dependent decay branching ratio to the Z or Higgs boson ($\tilde{t}_2 \to \tilde{t}_1 Z/H$). The hadronic mode of this decay again leads to multijet final states with little MET [$\tilde{t_2}\tilde{t_2^*} \to t\bar{t}(H/Z)(H/Z)$]. Our method may allow us to fully reconstruct the underlying new particles in such models, too.

5.2 Benchmark scenario: Ten-jet final state from vectorlike top pair production

In order to illustrate the strength of our jet tagging method, we investigate a simplified version of heavy particle production topology in this chapter, i.e. we consider the fully hadronic final state of pair-produced vectorlike top partners at the LHC ($pp \to T\bar{T} \to t\bar{t}HH$). For a theoretical motivation of this process also see Sec. 2.3. In particular, we study the performance of our taggers at the 14 TeV LHC with a vectorlike top of mass around 1 TeV as our benchmark scenario. Studies of fully hadronic final states in similar processes have been based on fat jet substructure [D], [144]–[146], including experimental searches at 8 TeV by CMS [147]. Current exclusion bounds on the vectorlike top mass are $m_T \gtrsim 700 - 950\,\text{GeV}$ from

ATLAS [148]–[150] and $m_T \gtrsim 690-910\,\text{GeV}$ from CMS [151]–[153], depending on the assumed branching fractions.

We extend the Standard Model Lagrangian by adding a vectorlike top T that interacts with the Standard Model top t and Higgs H,[1]

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \bar{T}(i\not{D} - m_T)T + y_T H\bar{t}T + h.c. \,. \tag{5.1}$$

We assume that the vectorlike top decays exclusively to a top and a Higgs. The mass of the vectorlike top in consideration is $m_T = 0.8 - 1.2$ TeV. The mass of the top is taken to be 173 GeV. The Standard Model Higgs has mass 126 GeV and decays to $b\bar{b}$ with branching ratio 56 %.

MadGraph5_aMC@NLO 2.2.1 [154] is used for generating parton-level events, which undergo hadronization and showering through Pythia 6.426 [155]. Delphes 3.1.2 [156], [157] with parameters tuned to the ATLAS detector is used for fast detector simulation.

The relevant Standard Model background processes for our analysis and their respective NLO K-factors are $pp \to t\bar{t}$ (1.61 [158]), $pp \to t\bar{t}b\bar{b}$ (1.77 [159]), $pp \to t\bar{t}H$ (1.10 [34]), and $pp \to bb\bar{b}\bar{b}$ (1.40 [160]). All final-state top quarks and Higgs bosons are decayed hadronically within MadGraph5_aMC@NLO. The following generator-level cuts are imposed: minimum transverse momentum of each outgoing parton $p_\perp \geq 20\,\text{GeV}$, angular separation between outgoing light quarks and between a light quark and a bottom quark $\Delta R_{jj}, \Delta R_{jb} \geq 0.2$, and angular separation between a pair of bottom quarks $\Delta R_{bb} \geq 0.4$. The latter cut is imposed to guarantee sufficient b separation to employ statistically independent b quark tagging. The overall scalar transverse momentum is imposed $H_T^{\text{parton level}} \geq 1\,\text{TeV}$, consistent with a similar (but stronger) cut at analysis level, cf. Eq. (5.9).[2] The cut on $H_T^{\text{parton level}}$ guarantees a reasonably large fraction of events in the signal regions. Note that this parton level cut on H_T makes it difficult to generate events at NLO, because it acts differently on processes with additional jets at matrix element level (the set of partons that contribute to the sum is different). Matching of matrix element with additional jets is also difficult for the same reason. Therefore, we generate background events at LO without matching to higher multiplicities at matrix element level. Thus, the absolute numbers of the background events should be taken with a grain of salt. The generated signal events do not suffer from this approximation.

5.3 Limits of substructure methods: Fat jet contamination

We briefly discuss the expected performance of algorithms using fat jets in the present process, $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets. We already argued in Sec. 5.1 that the fat jets are not well separated in such a busy hadronic final state, and this problem is illustrated in Fig. 5.1 for the signal process with vectorlike top mass $m_T = 1\,\text{TeV}$. The figure shows the angular distributions of the ten partonic (anti)quark final state (Monte Carlo truth) daughters. The black solid (dashed) line shows the distribution of the largest angular distance between the truth daughters of one top quark (Higgs boson) found in an event. The

[1] In general, there can also be a model-dependent term $\lambda H\bar{t}\gamma_5 T + h.c.$ in the Lagrangian. Here we assume $\lambda = 0$ for simplicity.

[2] To determine the respective cross-section at large scalar transverse momentum, we cut on events generated with $H_T^{\text{parton level}} \geq 500\,\text{GeV}$ to achieve better accuracy. Only for plotting we also generate $t\bar{t}$ and $bb\bar{b}\bar{b}$ events with $H_T^{\text{parton level}} \geq 1.2\,\text{TeV}$.

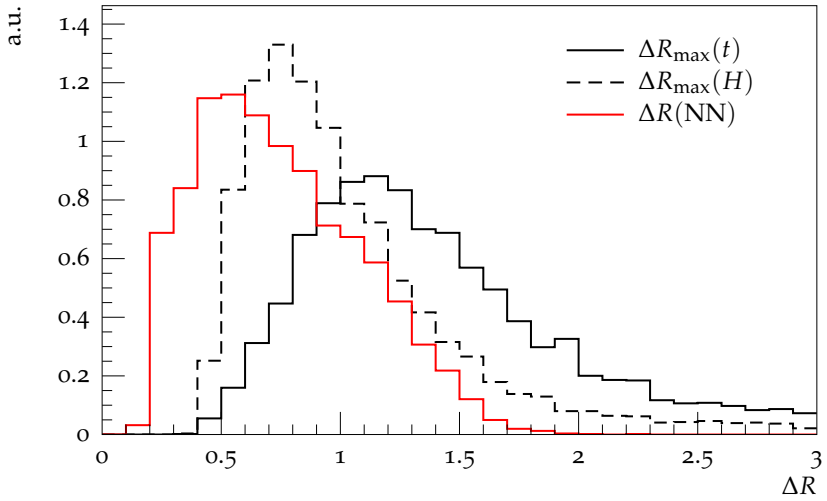


Figure 5.1: Angular distances between daughter particles from top and Higgs decays, in the benchmark process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets at the LHC with $\sqrt{s} = 14\,\text{TeV}$, with vectorlike top mass $m_T = 1\,\text{TeV}$ (parton level, arbitrary units). The largest R-distance between the daughters of the same top quark (Higgs boson) is denoted by $\Delta R_{\text{max}}(t)$ ($\Delta R_{\text{max}}(H)$) and plotted with black solid (dashed) lines. The minimal R-distance between the nearest neighbour daughters coming from different mothers $\Delta R(\text{NN})$ is depicted by the red line.

smallest distance between any truth daughters not coming from the same mother resonance is given by the red line. It is observed that the distance between the nearest daughter particles coming from different mother resonances, $\Delta R(\text{NN})$, is typically smaller than the angular spread of a t or H decay, $\Delta R_{\text{max}}(t/H)$. As a result, the fat jets will be contaminated.

To be more specific, we take the default fat jet clustering parameters of the widely used `HEPTopTagger` (cf. Sec. 3.6)

$$\text{Cambridge–Aachen}: \qquad R^{\text{fat jet}} = 1.5 \qquad \text{and} \qquad p_\perp^{\text{fat jet}} \geq 200\,\text{GeV}\,, \tag{5.2}$$

and give some concrete results for the process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets ($m_T = 1\,\text{TeV}$) in Fig. 5.2 (CA15, upper panels). A different choice of parameters is suggested in comparisons between boosted top tagging algorithms [19],

$$\text{anti-}k_T: \qquad R^{\text{fat jet}} = 1.0 \qquad \text{and} \qquad p_\perp^{\text{fat jet}} \geq 200\,\text{GeV}\,, \tag{5.3}$$

and we also give the plots for these fat jets in Fig. 5.2 (AKT10, lower panel).

The upper panels imply that the clustering radius of the CA15 jets is too large in this situation. The upper left panel shows the distribution of the number of fat jets. In more than 50% of events, only three fat jets are found, and even less in another 20% (Fig. 5.2 upper left). A fat jet is labelled "pure truth" t

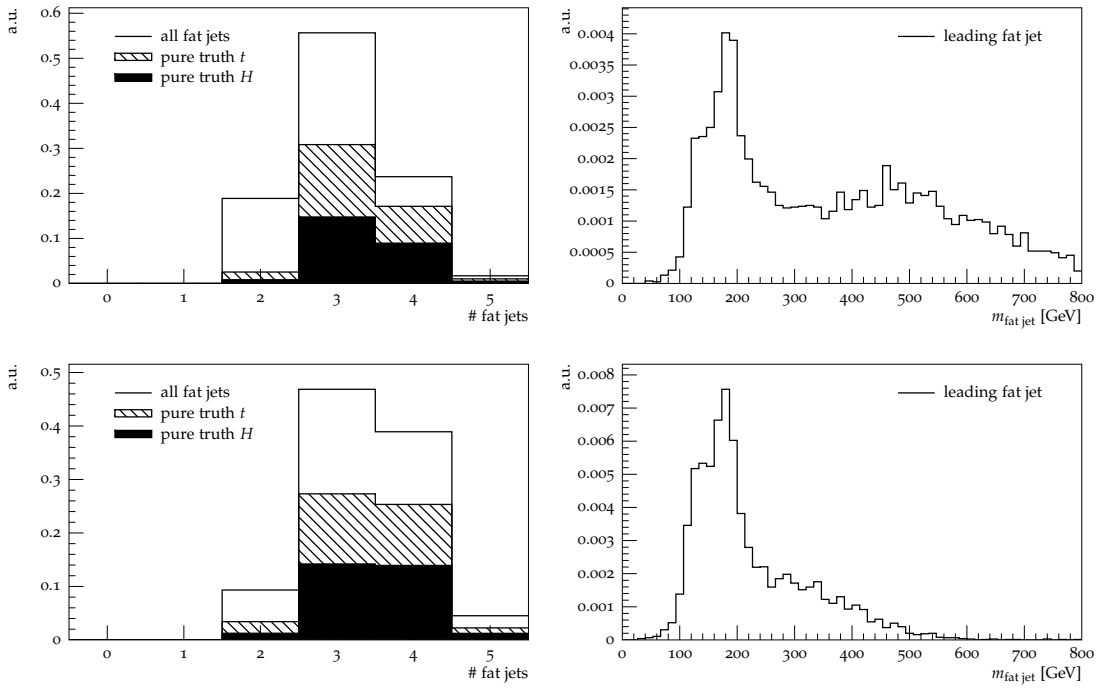


Figure 5.2: The upper row shows the results for CA15 fat jets: The number of fat jets is given in the left plot. A fat jet is labelled pure truth t (H) if all truth daughter partons of one top quark (Higgs boson) and no other truth daughter partons are ghost-associated. For events in which three fat jets were found, the distribution of the plain jet mass of the leading fat jet is shown in the right plot. Lower row: The same plots for AKT10 fat jets.

(H) if all truth daughter partons of one top quark (or Higgs boson) and no other truth daughter partons are ghost-associated [128].[3] In each bin, the fraction of pure truth t (H) fat jets are represented by the hatched (black) area. One can see that only a relatively small fraction of the fat jets are pure truth ones. A phenomenological study would have to rely on events with only three fat jets, and the plain jet mass of those leading fat jets is depicted in the upper right panel. There is a large tail toward very large jet masses, which suggests that there is a significant amount of splash-in from jets not coming from the same t or H resonance.

The second setup (AKT10) with a smaller-radius fat jets, shown in the lower row of Fig. 5.2, behaves better in this respect. In roughly 40% of events the correct number of four fat jets is identified (lower left), although three-jet events are still dominant. For the events with three fat jets, the leading jet mass is shown in the lower right panel. It can be seen that the distribution still shows a tail, but now large jet masses are much less present than in CA15 jets, implying less contamination through splash-in. On

[3]The truth partons' momenta are rescaled to infinitesimal $p_\perp$ and energy while η and ϕ are kept fixed ("ghosts"), and participate in jet clustering. Those partons that end up as constituents of a certain jet are called ghost-associated. Due to the vanishing energy of the ghosts, the final jets are unaffected.

the other hand, when we compare the fraction of pure jets in events with four fat jets between the two setups, we observe that AKT10 jets behave worse: While almost 75% of the respective CA15 fat jets are pure, this number is degraded to 65% for AKT10.

We conclude that the study of our high-multiplicity benchmark process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets is difficult if we rely on fat jets.[4] Also see Appendix 5.A at the end of this chapter for a numerical analysis using self-adaptive fat jets of variable size. The problem of insufficient separation of the boosted resonances (t and H) is not generically avoided even if a different fat jet radius is chosen – the smallest distance between truth daughters from different mothers is typically smaller than the angular spread of the top quark and Higgs boson, as shown in Fig. 5.1. There is no apparent solution to this contamination within fat jet algorithms, and the choice of clustering algorithm and parameters is related to finding a balance between splash-in (the fat jet contains energy deposit from a different resonance, too) and splash-out (the fat jet does not contain all radiation from a given resonance). As will be demonstrated in the following sections, this problem is circumvented when the mass-jump algorithm is used.

5.4 Back to separate jets – with a terminating veto

We investigate a new approach of analyzing high-multiplicity final states based on separately resolved jets. The process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets of Sec. 5.2 serves as the benchmark scenario. We try to answer the two key questions that arise in such an analysis: (1) which algorithm to use to construct the jets, and (2) how to reduce the sheer combinatorial choices of assigning the jets to the resonance particles of the process. We examine the first question by comparing the mass-jump algorithm of Chapter 4 with the corresponding jet clustering algorithm of the generalized-k_T family. The latter question is addressed by the bucket algorithm, which we review first.

5.4.1 Recap: The bucket algorithm

In high-multiplicity events, the assignment of jets to their respective heavy resonances can easily get out of hand. There are $6!/(3!3!) = 20$ possible combinations to assign six jets to two top quarks, and for eight jets coming from two tops and a Higgs boson this number already reaches $8!/(3!3!2!) = 560$. The bucket algorithm [140], [141] was proposed in the context of these two final states and introduces a "bucket" for each top quark (and one additional bucket of unassigned jets B_{ISR}), into which the jets are allocated. For a bucket B_i the metric

$$\Delta_{B_i} = |m_{B_i} - m_t| \qquad \text{with} \qquad m_{B_i}^2 = \left(\sum_{j \in B_i} p_j \right)^2 \tag{5.4}$$

measures the similarity of a collection of jets inside the bucket with a top quark. In Refs. [140], [141], the combination is determined by minimizing a global χ^2-like metric defined as

$$\Delta_{\text{global}}^2 = \omega \Delta_{B_1}^2 + \Delta_{B_2}^2 \,, \tag{5.5}$$

[4] For $m_T \lesssim 900\,\text{GeV}$, an analysis based on fat jets can still reconstruct the vectorlike top [D]. An experimental analysis of the same process also relies on fat jets [147].

and choosing a large $\omega = 100$ effectively decouples the two buckets. Thereby $\Delta_{B_1} < \Delta_{B_2}$ holds and the problem of unfeasible combinatorics is circumvented because the buckets can be filled independently. Top tagging is performed by imposing cuts on each bucket later on. Whereas in the original proposal the number of jets inside each bucket is not fixed, in our analysis we require strictly three jets in each top bucket, and also introduce Higgs buckets that contain exactly two jets.[5]

We apply the bucket algorithm in conjunction with mass-jump jet clustering as well as conventional jet clustering for comparison. Our benchmark scenario is given by $t\bar{t}HH$ production from a pair of vectorlike tops. Clearly the naive combinatorics are overwhelming even for the minimal final-state multiplicity of ten jets: $10!/(3!3!2!2!) = 25200$. To tackle this problem we formally define a global metric

$$\Delta^2_{\text{global}} = \omega_1 \Delta^2_{B_{t1}} + \omega_2 \Delta^2_{B_{t2}} + \omega_3 \Delta^2_{B_{H1}} + \omega_4 \Delta^2_{B_{H2}} \tag{5.6}$$

and explicitly decouple the four buckets by choosing the (positive) weights such that

$$\frac{\omega_{i+1}}{\omega_i} = +0 \quad i = 1..3\,. \tag{5.7}$$

Therefore, the buckets are filled separately in order $(B_{t1}, B_{t2}, B_{H1}, B_{H2})$ and the computational load is reduced to only $10!/7!/3! + 7!/4!/3! + 4!/2!/2! = 161$ comparisons.[6] In reality the number of jets will often be larger than the minimum of ten, where the speedup indicated here becomes even more prominent. A detailed description of our specific algorithm is given below. We address possible issues related to the explicit decoupling of the buckets in Section 5.5.2.

5.4.2 The analysis

We present an analysis that aims to identify the fully hadronic final state $t\bar{t}HH$ from vectorlike top pair production. We do not rely on large-radius "fat" jets and their substructure, which has become a standard approach whenever boosted heavy particles are involved. Conversely, the approach presented here focuses on separately resolved (small-radius) jets and is intended as a proof-of-concept in a realistic and relevant process.

The proposed analysis consists of the following steps, each of which is described in detail in the remainder of this subsection.

1. Event preselection cuts:
   Scalar transverse momentum $H_T \geq 1400\,\text{GeV}$ and number of b-tagged jets $\#b \geq 4$.

2. Jet reconstruction and cut #jets ≥ 10. Here, we use several different benchmark algorithms including the mass-jump algorithm.

3. Assignment of jets to the four buckets B_{t1}, B_{t2}, B_{H1}, and B_{H2} and cuts.

[5] In the top quark rest frame, in a large fraction of events, one of the decay products from $t \to bW^+ \to bjj$ carries low transverse momentum and thus fails to be reconstructed as a jet. As we are concerned with boosted top quarks from a heavy resonance decay, this problem does not occur.

[6] If the Higgs buckets are filled before the top buckets, the combination is further reduced to $10!/8!/2! + 8!/6!/2! + 6!/3!/3! = 93$. However, this would increase the wrong assignments for both the signal and background. See also discussion in Section 5.5.2.

4. Kinematic reconstruction of the vectorlike tops, depending on the number of identified top and Higgs buckets.

Event preselection

The decay cascade of a heavy vectorlike top pair leads to an energy deposit of $H_T \sim \mathcal{O}(2m_T)$ in the detector. H_T is the scalar transverse momentum, defined as

$$H_T \equiv \sum_{\text{jets } j} p_\perp^{(j)} \,. \tag{5.8}$$

We require

$$H_T \geq 1400\,\text{GeV} \tag{5.9}$$

to retain the majority of signal events for a vectorlike top with $m_T \sim 1\,\text{TeV}$ while strongly suppressing all nonresonant background processes. In the event preselection, jets are clustered with the anti-k_T algorithm as implemented in FastJet 3.0.6 with parameters $R = 0.4$ and $p_\perp \geq 20\,\text{GeV}$. Note that the jets are reconstructed differently after the preselection.

As the signal process contains six b quarks in the final state, we also cut on the number of bottom tags. Tagging is performed by Delphes using the jets defined above. We select a conservative working point where 70% of b-initiated jets are identified correctly, $\epsilon_{\text{tag}} = 0.70$, and assume the mistag rates of charm-initiated jets to be $\epsilon_{\text{mis}}^{(c)} = 0.10$, and $\epsilon_{\text{mis}}^{(udsg)} = 0.01$ for light (quark- or gluon-initiated) jets. Cutting on

$$\#b \geq 4 \tag{5.10}$$

reduces the relevant backgrounds to b-rich processes with high multiplicity, $pp \to t\bar{t}$, $pp \to t\bar{t}b\bar{b}$, $pp \to b\bar{b}b\bar{b}$, and $pp \to t\bar{t}H$. In particular we estimate the rate of mistagged bottom quark pair production to be less significant than the $b\bar{b}b\bar{b}$ process. By the same argument we can neglect pure QCD multijet production, which cannot be simulated reliably and is typically extracted from data.

Jet reconstruction

Jets are reconstructed from all calorimeter towers that lie within $|\eta| < 4.9$. To avoid "chopped" jets at the boundary of the detector, we require $|\eta_{\text{jet}}| < 4.0$ so that all jets are sufficiently central. The key ingredient to this analysis is the choice of jet clustering algorithm. In our study, we adopt the following benchmark algorithms and compare them:

- A C/A-like mass-jump clustering algorithm with parameters

$$[\text{MJ06}] : (\, R = 0.6\,, \quad p_\perp \geq 25\,\text{GeV}\,, \quad \theta = 0.7\,, \quad \mu = 50\,\text{GeV}\,) \,. \tag{5.11}$$

$$[\text{MJ10}] : (\, R = 1.0\,, \quad p_\perp \geq 25\,\text{GeV}\,, \quad \theta = 0.7\,, \quad \mu = 50\,\text{GeV}\,) \,. \tag{5.12}$$

- A standard setup with the Cambridge-Aachen algorithm and commonly used clustering parameters

$$[\text{CA03}] : (\, R = 0.3\,, \quad p_\perp \geq 25\,\text{GeV}\,) \,. \tag{5.13}$$

$$[\text{CA04}] : (\, R = 0.4\,, \quad p_\perp \geq 25\,\text{GeV}\,) \,. \tag{5.14}$$

The minimum jet $p_\perp$ is set to be the same to allow for easy comparison of the results. The additional veto parameters specific to mass-jump clustering, θ and μ in Eqs. (5.11) and (5.12), are motivated by the results obtained from boosted top quarks, cf. Sec. 4.3.2. The mass-jump veto leads to jets whose effective radius can vary, and it is this inherent flexibility that will lead to improved results compared to standard jet clustering with fixed angular size.

Because some jets reconstructed with the mass-jump algorithm may experience a very large effective radius,[7] contamination from pile-up and underlying event can pose problems in a realistic environment. We therefore apply a trimming stage (cf. Sec. 3.3) with parameters $R_{\text{trim}} = 0.2$ and $f_{\text{trim}} = 0.03$ as suggested in the original paper [94]. Trimming is applied to all the benchmark points, MJ06, MJ10, CA03, and CA04.

After the jets are reconstructed, we require

$$\#\text{jets} \geq 10\,, \tag{5.15}$$

three for each top quark and two for each Higgs boson.

Bucket construction and tagging

In order to keep the combinatorial choices of this multijet process at a manageable level, we make use of the idea of buckets as reviewed above. First of all, the first top bucket B_{t1} is filled with the three jets that minimize

$$\Delta = |m_{\text{bucket}} - m_t|\,. \tag{5.16}$$

Here and for all other buckets, we limit the allowed jet combinations to those that fulfill

$$p_{\perp,\text{bucket}} \geq 200\,\text{GeV}\,. \tag{5.17}$$

This prevents wrong assignments from widely separated low-energy jets, which are possible due to the sheer number of possible choices. In addition, only combinations with minimum mutual jet separation

$$\Delta R(j,j) \geq 0.3 \tag{5.18}$$

are considered, because smaller distances cannot reasonably be resolved by the hadronic calorimeter any more. This cut is also consistent with the cuts applied on generator level, cf. Sec. 5.2. Note that we do not impose an upper cut on angular spread of the top decay products, as is implicitly done in all substructure methods which rely on a fat jet of fixed radius. Also note that Eq. (5.18) does not restrict the analysis if all jets are mutually separated by more than $\Delta R(j,j) = 0.3$, i.e. the fixed-R setups CA03 and CA04 are unaffected. If two top subjets are very close by and merge in the CA03 setup, even in the ideal case that the MJ algorithm can resolve them separately, they could not contribute to the same bucket. In this sense the cut helps to allow a fair comparison between the mass-jump setups and the Cambridge-Aachen setups.

After the first top bucket has been fixed, out of the remaining jets the second top bucket B_{t2} is filled with three jets, then the first Higgs bucket B_{H1} with two jets, and finally B_{H2} again with two jets. This

[7] This effect will be investigated later, cf. Fig. 5.5.

	SR1	SR2	SR3
number of tagged t	$= 1$	$= 2$	$= 2$
number of tagged H	$= 2$	$= 1$	$= 2$

Table 5.1: The three signal regions.

course of action corresponds to a global metric with explicitly decoupled buckets as defined in Eqs. (5.6) and (5.7). Again for each bucket, out of all possible jet combinations that fulfill Eqs. (5.17) and (5.18), the combination with minimum metric (Eq. (5.16), where m_t is replaced by m_H for Higgs buckets) is selected. Events where there is no viable jet combination for a bucket are negligibly rare. Remaining jets are assigned to a fifth bucket B_{ISR} and not further considered in our analysis.

Only after all buckets have been filled, cuts are applied. For top candidate buckets, we require

$$\Delta \leq 25\,\text{GeV}\,, \tag{5.19}$$

$$\left(\frac{m_W}{m_t}\right)_{\text{bucket}} \in \frac{m_W}{m_t} \pm 15\%\,, \tag{5.20}$$

$$\left(\frac{m_{23}}{m_t}\right)_{\text{bucket}} \geq 0.35\,. \tag{5.21}$$

Eq. (5.19) is a simple cut on the reconstructed top mass. The mass ratio in the left-hand side of Eq. (5.20) is constructed from the two jets that best reconstruct the W boson mass ($m_{W,\text{bucket}}$) and the total jet mass of the bucket ($m_{t,\text{bucket}}$), as proposed when the bucket algorithm was introduced in Ref. [140]. The final cut in Eq. (5.21) was introduced in the `HEPTopTagger`, where m_{23} is the combined mass of the two subleading jets in the bucket (in terms of $p_\perp$). In our study, it helps to suppress top candidates whose momentum is dominated by one very hard prong.

Higgs candidate buckets have to fulfill

$$\Delta \leq 20\,\text{GeV}\,. \tag{5.22}$$

The 4-momentum of the successful top or Higgs candidate is given by the momentum sum of the jets inside the bucket.

Signal regions and kinematic reconstruction

We define three signal regions depending on the number of tagged buckets, see Tab. 5.1. In addition to event rates, we kinematically reconstruct the vectorlike top from the momenta of a tagged top quark and a Higgs boson to assess its invariant mass

$$M(t,H) = \sqrt{(p_t + p_H)^2}\,. \tag{5.23}$$

In the case of a fully reconstructed event (SR3), we choose between the two possible pairings, $\{(t_1,H_1),(t_2,H_2)\}$ and $\{(t_1,H_2),(t_2,H_1)\}$, such that the mass difference of the two vectorlike tops is minimal,

$$\min\left[|M(t_1,H_1) - M(t_2,H_2)|, |M(t_1,H_2) - M(t_2,H_1)|\right]\,. \tag{5.24}$$

Process		$T\bar{T}$					b.g.	$bb\bar{b}\bar{b}$	$t\bar{t}$	$t\bar{t}b\bar{b}$	$t\bar{t}H$
		800 GeV	900 GeV	1.0 TeV	1.1 TeV	1.2 TeV					
number of events for $100\,\text{fb}^{-1}$											
$H_T \geq 1.4\,\text{TeV}$		507	306	167	86.7	43.6	25600	4130	20600	772	52.9
$\#b \geq 4$		356	217	118	60.8	30.6	1730	990	506	218	16.4
MJ10	$\#\text{jets} \geq 10$	306	185	101	52.7	26.8	518	166	201	141	9.5
	SR1	14.9	10.4	6.4	3.8	1.9	3.3	0.5	1.1	1.5	0.1
	SR2	36.5	20.7	11.5	5.7	2.7	22.2	1.8	11.4	8.1	0.9
	SR3	10.5	6.0	3.9	2.2	1.0	2.1	0.1	1.1	0.8	0.1
CA03	$\#\text{jets} \geq 10$	282	172	90.6	45.6	22.9	392	121	145	118	7.7
	SR1	8.4	8.0	4.7	2.4	1.2	2.4	0.4	1.0	0.9	0.1
	SR2	24.4	14.1	6.8	3.6	1.8	11.1	0.8	5.3	4.5	0.5
	SR3	5.6	2.9	1.8	1.1	0.4	0.8	0.0	0.4	0.4	0.1

Table 5.2: Expected event numbers for two benchmark setups, mass-jump clustering MJ10 [Eq. (5.12)] and standard Cambridge-Aachen clustering CA03 [Eq. (5.13)]. Numbers are given for an integrated luminosity of $100\,\text{fb}^{-1}$ at the LHC with $\sqrt{s} = 14\,\text{TeV}$. Results for the signal are shown separately for different values of the vectorlike top mass ranging from 800 GeV to 1.2 TeV. All relevant background processes as well as their sum ("b.g.") are given in the right-hand columns. The three signal regions (SR) are defined in Table 5.1.

The majority of events, however, fall into signal regions 1 and 2, and we are left with three tagged and one untagged bucket. As the untagged bucket also contains a significant energy deposit, its momentum can be used as an estimate for the fourth particle. Again we apply Eq. (5.24) to determine the correct pairing. Only the vectorlike top that is reconstructed from two tagged buckets is further considered.

5.4.3 Results

The cut flow and expected event numbers at the LHC14 with $100\,\text{fb}^{-1}$ integrated luminosity are shown in Tab. 5.2 for two benchmark setups MJ10 [Eq. (5.12)] and CA03 [Eq. (5.13)]. For both setups, the $T\bar{T}$ signal outnumbers the Standard Model background up to $m_T = 800 - 900\,\text{GeV}$ in signal region SR2, and up to $m_T = 1.1\,\text{TeV}$ in the signal regions SR1 and SR3. Note that the absolute numbers should be taken with great care due to the simplified event generation setup, cf. Sec. 5.2. A comparison of the relative significances between the employed algorithms is less affected by the uncertainties, though.

We observe that event numbers are largest in SR2 (2 tagged top quarks, 1 tagged Higgs boson). This is particularly pronounced for the various background processes. It can be understood by the order in which the four buckets are filled: The top buckets are filled first and reconstruct the truth partons very well, as will be investigated in Sec. 5.5. If jets originating from a Higgs boson are wrongly assigned to a top bucket, it becomes unlikely to fill both Higgs buckets from the remaining jets with masses within

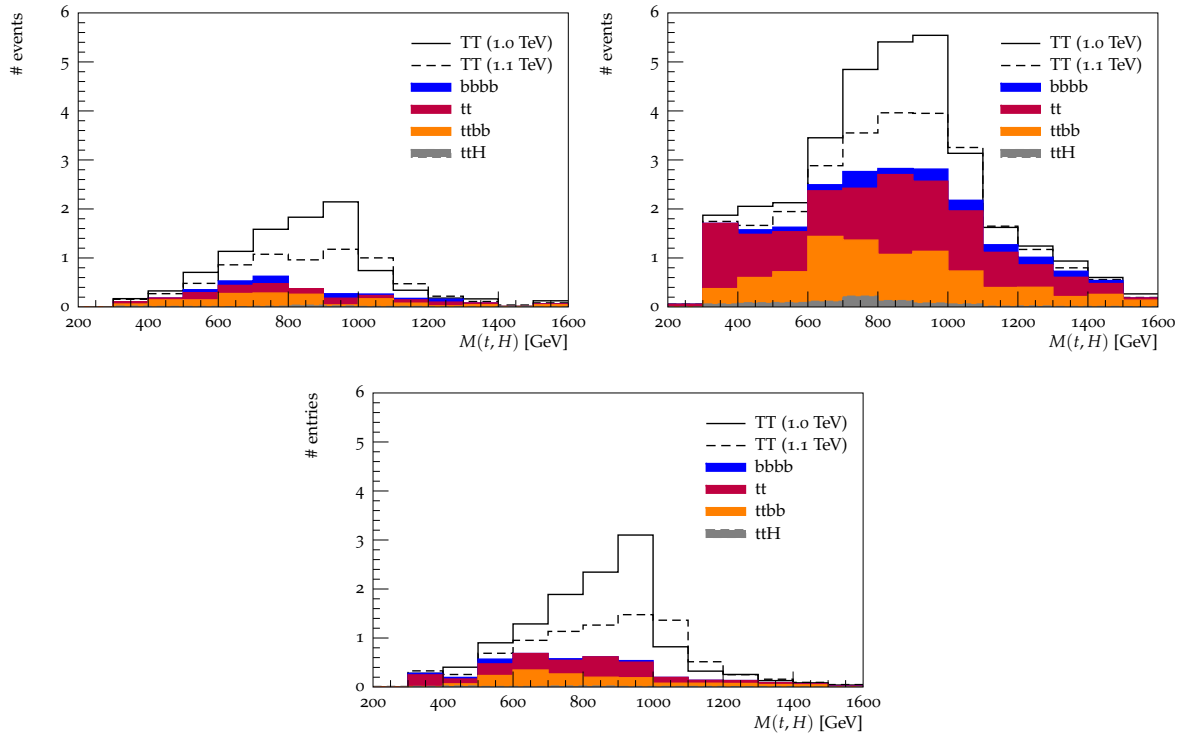


Figure 5.3: Reconstructed mass of the vectorlike top $M(t,H)$ (for truth $m_T = 1\,\text{TeV}$ and $1.1\,\text{TeV}$) with the MJ10 setup for SR1 (top left), SR2 (top right) and SR3 (bottom), for an integrated luminosity $100\,\text{fb}^{-1}$. The histograms are stacked.

the mass window. This effect is larger for background processes, among which only a vanishing fraction contains actual Higgs bosons at parton level except the $t\bar{t}H$ background. Thus the Higgs buckets are dominantly filled from the remaining unrelated jets.

As can be seen in Tab. 5.2, in the conventional clustering setup CA03 event numbers are considerably smaller than those obtained with mass-jump clustering MJ10, both for the signal and Standard Model backgrounds. This is already observable at the #jets ≥ 10 cut stage, and the difference becomes even larger when events in the final signal regions are compared. Due to the fixed jet radius of CA03, hard prongs that are separated by a distance smaller than $R = 0.3$ merge. It is thus easily understood that the number of hard jets is naturally smaller than the one obtained from a (reasonable) mass-jump setup. As our implementation of the bucket algorithm explicitly requires resolved constituent jets, those merged jets fail to reconstruct their hard resonance, leading to a large drop in event numbers in all signal regions.

In Fig. 5.3, we show the distributions of the vectorlike top mass, where stacked histograms of all three signal regions SR1 – SR3 are presented. (In SR3, each event gives two entries.) The kinematic reconstruction of the vectorlike top works very well, as manifest in a clear peak in the figures.

In order to compare the different jet clustering setups, it is instructive to look at signal significance

$S/\sqrt{B}$	800 GeV	900 GeV	1.0 TeV	1.1 TeV	1.2 TeV
MJ10	11.77	7.05	4.13	2.22	1.07
MJ06	11.38	6.96	4.06	2.16	1.02
CA03	10.17	6.63	3.49	1.86	0.90
CA04	11.06	5.91	3.36	1.51	0.61

Table 5.3: Comparison of significance $S/\sqrt{B}$ (number of signal events S and number of background events B summed over all three signal regions) for different jet algorithms and benchmark setups for $100\,\text{fb}^{-1}$ integrated luminosity. The numbers for the MJ10 and CA03 setups can be calculated from Tab. 5.2.

$S/\sqrt{B}$, which we take from the number of signal events S and number of background events B summed over all three signal regions. Numbers are given in Tab. 5.3 for all considered setups. It is observed that among the standard clustering setups CA03 and CA04, the smaller jet radius yields better results. The reason is that nearby prongs can only be separately resolved if the radius parameter is smaller than the mutual separation. For mass-jump clustering MJ10 and MJ06, the opposite is true: the larger maximum jet radius gives more significant results. Even in very busy final states some prongs end up fairly isolated, and they are more accurately reconstructed with larger jets. Overall the mass-jump algorithm outperforms the fixed-radius conventional clustering.

The reconstructed signal mass (for truth $m_T = 1\,\text{TeV}$) is shown in Fig. 5.4 for all setups. A peak is visible for all jet clustering setups, but for the fixed-radius algorithms CA03 and CA04 it is shifted to lower values in the SR1 and SR2. The reconstruction is worse for the CA03 setup. Only the analysis based on the mass-jump clustering can reproduce the mass of the heavy T in all signal regions. Independent of the specific clustering algorithm, the reconstructed mass peak has an edge around the true mass, with the majority of events experiencing a lower value. This may be due to the fact that we do not explicitly include the leading gluon emission when the buckets are reconstructed.

Possible explanations for these observations and a comparison between standard Cambridge-Aachen and the mass-jump jets are given in the following subsection.

5.4.4 Comparison of jet clustering algorithms

The results found in the previous subsection have mixed implications for the ideal jet radius when standard fixed-R clustering is employed. CA03 yields larger overall significance than CA04, cf. Tab. 5.3. This is not surprising, as only a small radius can separately resolve hard prongs from boosted top and Higgs decays. In terms of event numbers, this advantage seems to well compensate for possible splash-out, a loss of final-state radiation that falls outside the cone. On the other hand, Fig. 5.4 shows that the reconstruction of the vectorlike top mass works better with a larger radius. This illustrates the difficulty to find an optimal radius R in the fixed-R clustering algorithms.

Instead of employing a fixed clustering radius, the mass-jump algorithm was designed to separately resolve hard prongs at any distance scale if the terminating veto is called. Fig. 5.5 shows the angular distance $\Delta R = \sqrt{d_{ij}}$ of the last recombination step in the hardest (left) and tenth-hardest jet (right). Whereas for CA03 jets (dashed lines) the ΔR distribution peaks at the radius cut $R = 0.3$ or slightly

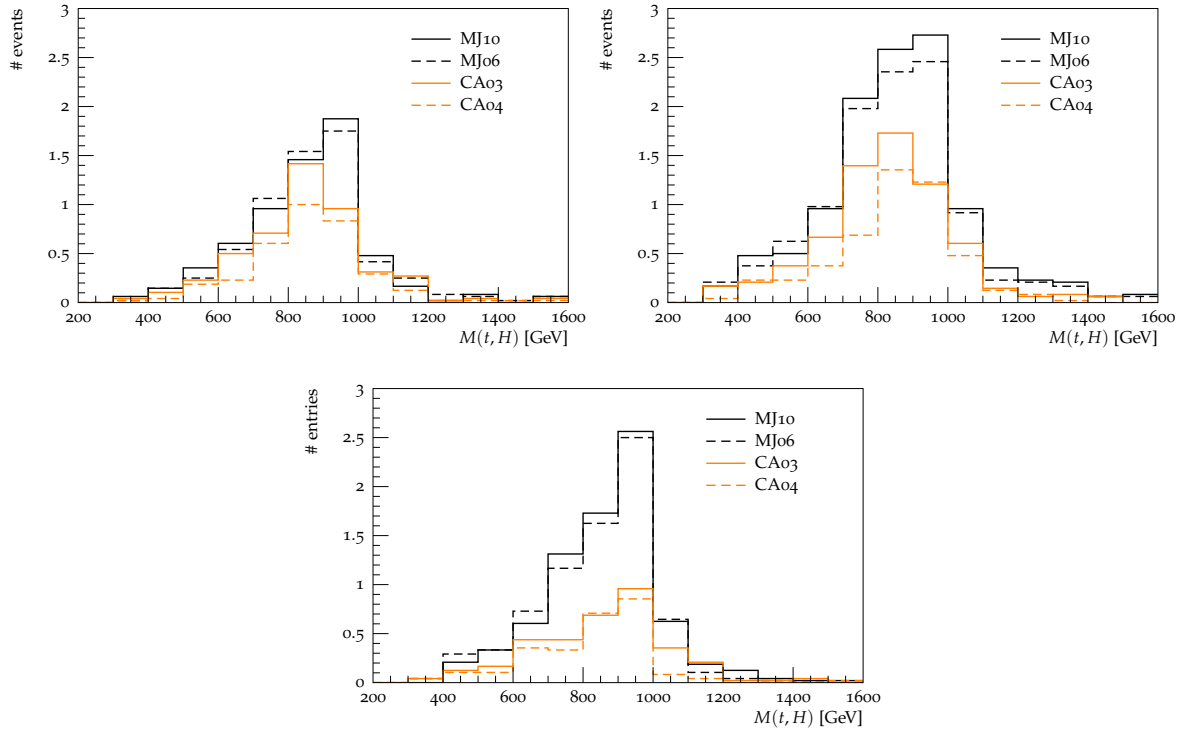


Figure 5.4: Reconstructed mass of the vectorlike top (for truth $m_T = 1\,\text{TeV}$) in SR1 (top left), SR2 (top right) and SR3 (bottom) for different jet algorithms, for an integrated luminosity $100\,\text{fb}^{-1}$.

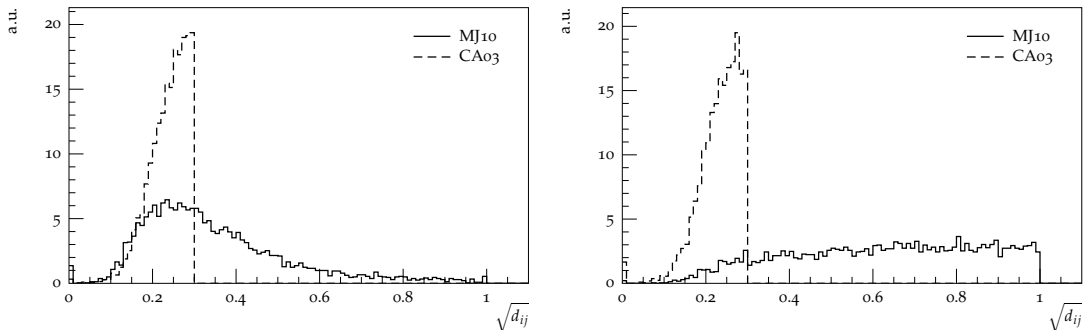


Figure 5.5: $\Delta R = \sqrt{d_{ij}}$ of the last recombination step in the hardest (left) and tenth-hardest jet (right) in signal events with $m_T = 1\,\text{TeV}$ (arbitrary units). The solid lines depict values for jets clustered with the C/A-like mass-jump algorithm (MJ10), whereas jets clustered with the conventional C/A algorithm (CA03) are given by dashed lines.

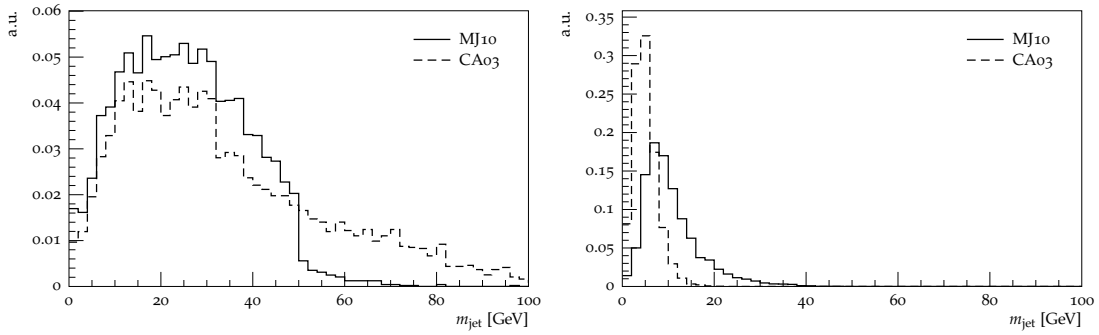


Figure 5.6: Trimmed jet mass of the hardest (left) and tenth-hardest jet (right) in signal events with $m_T = 1\,\text{TeV}$ (arbitrary units). The solid lines depict values for jets clustered with the C/A-like mass-jump algorithm (MJ10), whereas jets clustered with the conventional C/A algorithm (CA03) are given by dashed lines.

below, MJ10 jets (solid lines) observe much more variety. For the hardest jet, ΔR has a peak at around 0.25 but can also take a large value, and for the tenth-hardest jet it has a broader and almost flat distribution. This inherent flexibility constitutes the key to reconstructing the busy final state considered here.

The tail up to very large values in ΔR seen for the mass-jump tenth (soft) jets (Fig. 5.5 r.h.s.) may be a relic of the algorithm. In terms of significance, it was nevertheless observed that the overall performance is improved when such large radii are allowed.[8] These large-area jets can gather additional soft radiation, e.g. soft gluon emissions, which can lead to a more accurate bucket mass. The fixed-R setups CA03 and CA04 do not have these features, which could explain why the reconstructed T mass in SR1 and SR2 is shifted to lower values for these algorithms. We speculate that a dedicated study of this effect may lead to improved taggers in this context, but it is beyond the scope of this work.

Fig. 5.6 shows the trimmed jet mass (cf. Sec. 3.3), again for the hardest (left) and tenth-hardest jet (right). A fraction of events experiences a very heavy leading jet around $m_j = 70 \sim 80\,\text{GeV}$ in the CA03 setup, indicating that nearby hard prongs have merged. The leading mass-jump jet, on the other hand, has a cutoff at $m_j = \mu = 50\,\text{GeV}$ due to the veto condition (cf. Sec. 4.2.2), and very large jet masses are absent. As plain jet mass roughly scales with $p_\perp \cdot R$, soft jets clustered with fixed-R algorithms tend to be very light, as shown in the right panel of Fig. 5.6 for the CA03 setup. However, final-state radiation of low-$p_\perp$ jets is less collimated and ideally caught in jets with larger radius [135]. In the same figure, it is seen that the tenth (soft) MJ10 jets are heavier due to the larger effective jet radius as observed in Fig. 5.5.

[8] This improvement is diminished due to trimming. By including a trimming stage (cf. Sec. 3.3), we assume that our results are not affected much if additional soft radiation from underlying event and pile-up are taken into account. These effects should be included in a realistic study, but pile-up can only be reliably simulated by the experimental collaborations. We assume that our results, in particular the comparison between conventional jet clustering and mass-jump clustering, are still qualitatively valid in our simplified setup.

We conclude that the results found in Sec. 5.4.3, namely that a small jet radius can be of advantage for conventional fixed-R clustering algorithms, whereas mass-jump clustering benefits from a very large maximum R, can be explained by looking at jet merging scales and mass distributions. For our process including four boosted resonances and a very busy final state, it is essential to find jets with a flexible algorithm. The mass-jump algorithm avoids the problem of searching for a good compromise for the fixed jet radius parameter and leads to physically more appealing jets. Consequently, it generally outperforms its standard fixed-R counterpart, the Cambridge-Aachen algorithm, in the phenomenological analysis.

5.5 Performance of top and Higgs tagging

In this section, we investigate the performance of top and Higgs tagging in our approach with the mass-jump (and the Cambridge-Aachen) clustering algorithms. We also briefly comment on the metric of the decoupled buckets in Sec. 5.5.2.

It should be emphasized that the tagging efficiencies and the quality of reconstruction depend on the considered physical processes as well as the event generator by which the samples have been produced. This is even more true in our study, as the performance of top/Higgs tagging is affected by the hadronic activity from other top and Higgs decay products in the candidate's vicinity. Tagging efficiencies and reconstruction qualities of the present canonical tagging algorithms for boosted resonances are usually evaluated for isolated fat jets (see e.g. Refs. [19], [21]), which reduces the dependence on the specific process and makes it possible to compare the results between different algorithms. This condition is not satisfied in our benchmark analysis and therefore the results for top and Higgs tagging can hardly be related to other algorithms. In addition, the strong weighting of the global buckets metric in Eq. (5.6) naturally leads to the first top bucket being much better reconstructed than the second. Similarly, the reconstruction quality of the top quarks is generally better than that of the Higgs bosons.

Despite those precautionary warnings, the results presented here can serve as a benchmark for other processes with a similarly busy final state.

5.5.1 Reconstruction quality

In Fig. 5.7 we assess the quality of momentum reconstruction of the tagged buckets for the preferred MJ10 setup, in the benchmark process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets at the LHC with $\sqrt{s} = 14\,\text{TeV}$. The reconstructed masses of the top quarks and Higgs bosons are shown in the upper row. Due to the ordering in the metric, the first bucket always gives a better reconstructed mass, leading to the dip for the second bucket. Both for top and Higgs candidates, there is a central peak at the true mass value for the first buckets. The top mass peak is much narrower, which is not surprising since Higgs buckets are filled by the remaining jets only after the two top buckets have been filled. The middle and lower rows of Fig. 5.7 show the deviation between the bucket momenta and the MC truth parton momenta in terms of two variables,

$$\frac{\Delta p_\perp}{p_\perp^{\text{reco}}} \equiv \frac{p_\perp^{\text{reco}} - p_\perp^{\text{parton}}}{p_\perp^{\text{reco}}} \qquad \text{and} \tag{5.25}$$

$$\Delta R_{\text{reco,parton}}\,. \tag{5.26}$$

A narrow peak around 0 is observed for both observables, for top as well as Higgs candidates. In fact, most of the tagged buckets are reconstructed within 20% in $\Delta p_\perp / p_\perp^{\text{reco}}$, and $\Delta R \leq 0.1$, with no significant differences between the respective first and second buckets. We conclude that (i) the tagged buckets are built from the correct jets, and that (ii) these jets reconstruct the truth partons' momenta very well.

For completeness, we also show the same results for the standard clustering benchmark setup CA03 in Fig. 5.8. Note that, although the distributions look similar to the MJ10 setup, the total number of tagged buckets is significantly smaller. The reconstructed Higgs bosons tend to have a broader peak, shifted to lower values in the CA03 setup. As a result, the mass of the reconstructed vectorlike top is also shifted to lower values, as has been observed in Fig. 5.4. This is another indication that the mass-jump algorithm is better suited to this analysis, in addition to larger event numbers discussed in Sec. 5.4.3.

We conclude our discussion with the reconstruction quality of tagged top buckets from the dominating $t\bar{t}$ Standard Model background (in the MJ10 setup), which is shown in Fig. 5.9. Deviations from the MC truth partons are very small and our analysis setup is well suited for the background processes as well. We observe that, although transverse momentum is reconstructed generally very accurately, the buckets tend to have lower values than the signal case (cf. Fig. 5.7). Final-state radiation off the boosted top quark may escape from the respective top bucket, and additional hard prongs from the matrix element that could lead to splash-in are not present.

A final comment about the accurateness of the reconstruction of Higgs bosons is in order. Although the buckets show very good quality (see again Fig. 5.7), the reader might wonder about the absence of a clear structure in the found mass. As noted above, this feature is to some extend expected because the Higgs buckets are filled only after the top buckets. In Fig. 5.10 we demonstrate that our setup in general does not affect the reconstruction of Higgs bosons adversely. From left to right, the plots show the reconstructed mass and the radial distance to the MC truth Higgs in a modified analysis where the Higgs buckets are built first. Note that due to the multijet final state there is a large number of wrong assignments compared to our preferred setup. As a result, the complete analysis gives a significant number of events only in signal region 1 and suffers from large backgrounds. Fig. 5.10 right-hand side shows the mass of all those Higgs buckets reconstructed with good quality, defined by $\Delta p_\perp / p_\perp^{\text{reco}} < 0.2$ and $\Delta R \leq 0.5$. The distribution is not biased by the mass-jump jet algorithm or our simulation.

5.5.2 A note on the global metric

The reader might wonder whether the jets are not optimally assigned to the buckets due to the explicitly decoupled metric in Eqs. (5.6) and (5.7). This choice was made to reduce the combinatorial workload, but it is expected that the results do not change much even if a more democratic ordered metric, $0 < \omega_{i+1}/\omega_i < 1$, is used. First, we observe that for any bucket the exchange of a jet with one from B_{ISR} does not yield a lower metric by definition, independent of its weights. Second, interchange of jets between two buckets (B_i and B_j with $\omega_i > \omega_j$) may lower the measure of B_j, but always at the cost of raising that of B_i. Because of the relative weight $\omega_i > \omega_j$, most of the interchanges are likely to increase the global measure. To find the global minimum, one has to consider a reassignment of several jets simultaneously, the details of which depend on the specific weights chosen and is beyond the scope of this analysis. Note that, even if finite weights are used, the local minimum found with explicitly decoupled buckets gives an

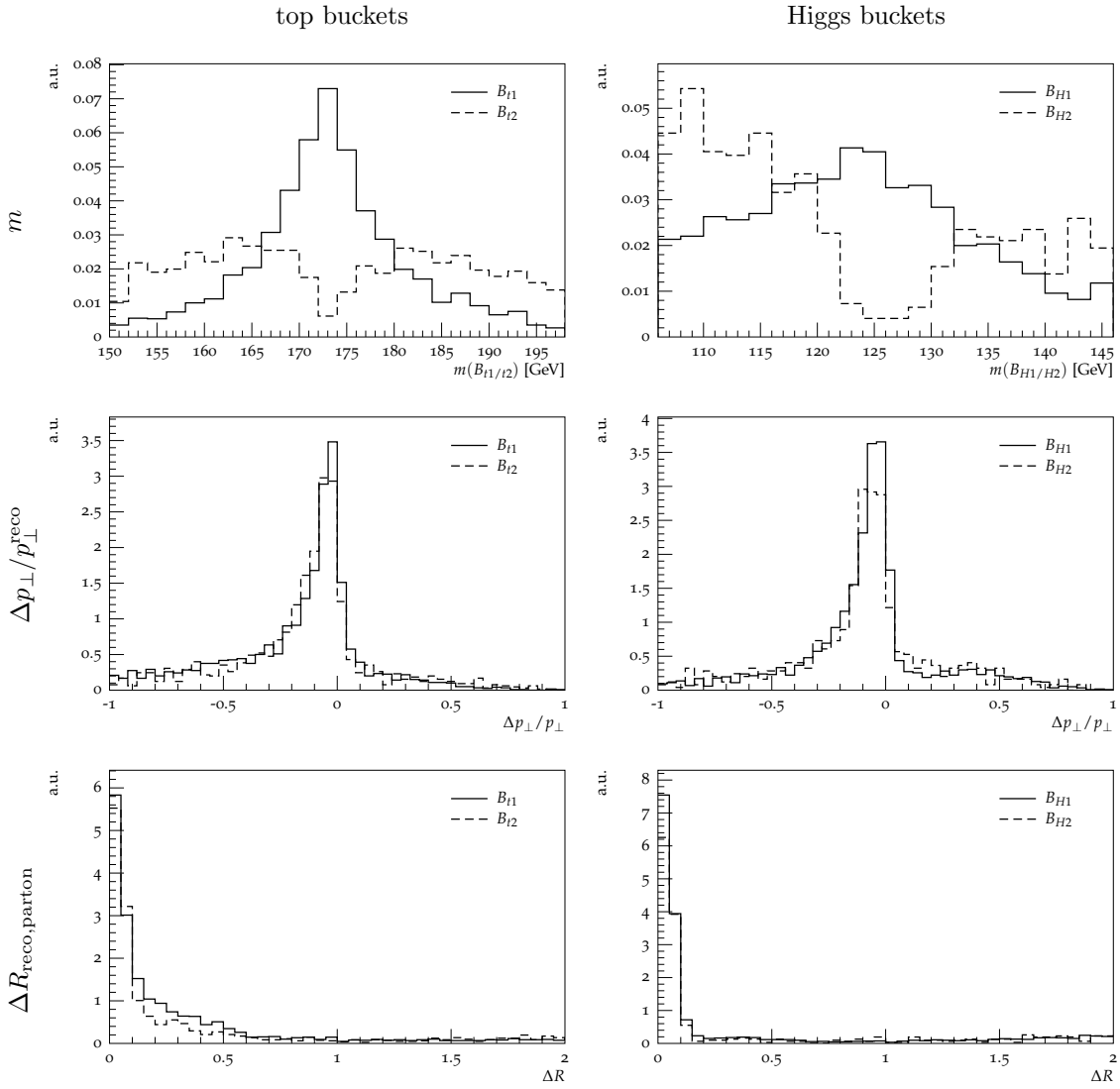


Figure 5.7: Reconstruction quality of tagged top (left) and Higgs buckets (right) for the MJ10 setup, in the benchmark process $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets at the LHC with $\sqrt{s} = 14\,\mathrm{TeV}$ (arbitrary units). From top to bottom, the reconstructed mass, relative deviation in transverse momentum $\Delta p_\perp / p_\perp^{\mathrm{reco}}$, and the angular distance $\Delta R_{\mathrm{reco,parton}}$ are shown. The solid curves show results for the first bucket, the dashed curves for the second bucket.

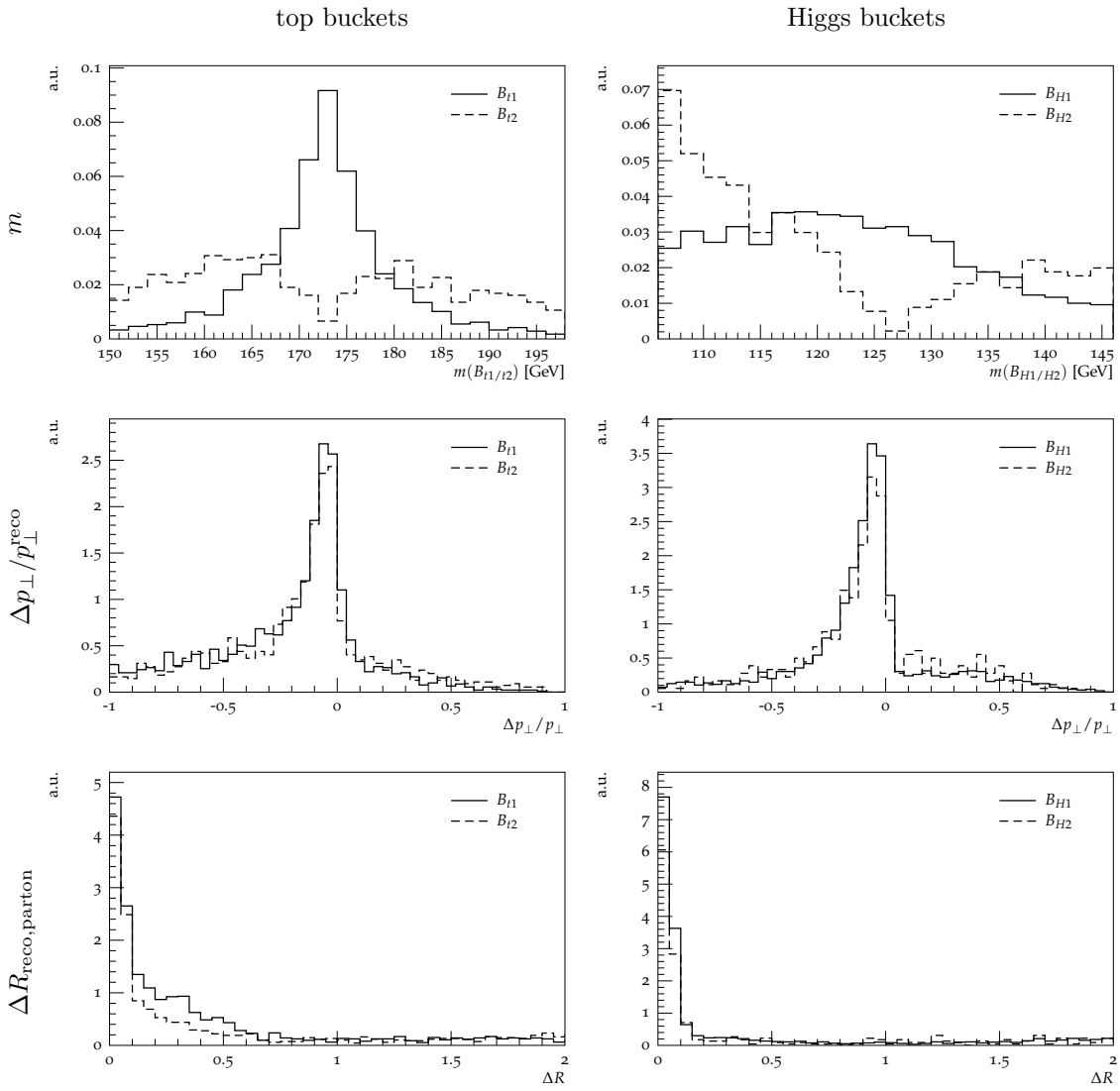


Figure 5.8: The same as Fig. 5.7 but for the CA03 setup.

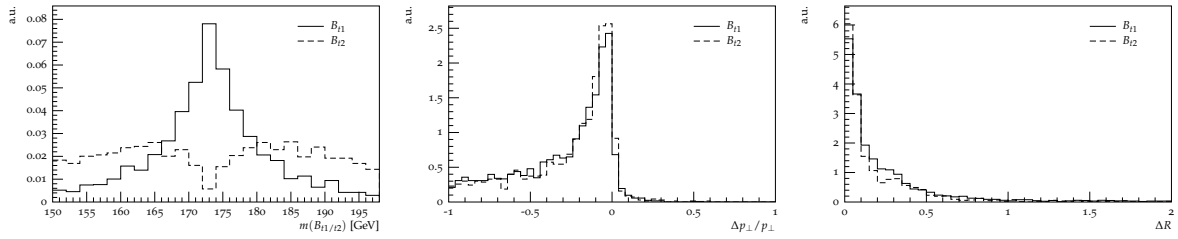


Figure 5.9: Reconstruction quality of tagged top buckets of the leading $t\bar{t}$ background events in the MJ10 setup.

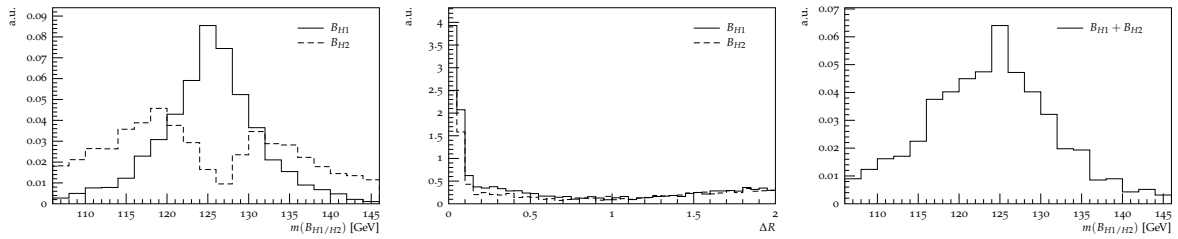


Figure 5.10: Reconstruction quality of tagged Higgs buckets for the MJ10 setup in a modified analysis where the Higgs buckets are filled first. There is a clear peak in the reconstructed mass (left), but a large number of buckets does not correspond to a MC truth Higgs boson (middle) compared to Fig. 5.7. Taking only well-reconstrued buckets into account (see definition in the text), the mass reconstruction still exhibits a central peak around the true mass value (right). We do not distinguish between the two buckets here.

upper bound on Δ^2_{min}, thus helping to reduce the huge number of permutations.

This argument is weakened if a metric is chosen that does not favour top buckets over Higgs buckets, i.e. $\omega_1, \omega_2 \not> \omega_3, \omega_4$. In our analysis, we chose to reconstruct top quarks from three prongs first, and only after that Higgs bosons from two prongs each. This order reduces wrong assignments for both the signal and background. Since Standard Model processes containing Higgs bosons in the final state are rare, the mass distributions in the Higgs buckets can serve as side bands to experimentally determine the background cross-sections.

5.6 Summary and outlook

We have presented a novel approach to the very busy all-hadronic final state emerging from multiple heavy resonances. We focused on vectorlike top pair production at the LHC as the benchmark process of our studies. Since the standard techniques using large-radius fat jets suffer from splash-in contamination and jet overlap in such a busy environment, here we completely relied on separately resolved jets. It was shown that this approach – in combination with a bucket algorithm to reduce computational weight – gives good results and can serve as an alternative channel in new physics searches, including a kinematic reconstruction of the vectorlike top mass. The key ingredient is the mass-jump jet clustering algorithm, which is shown to greatly improve the performance compared to common jet algorithms. This algorithm, which established the family of jet clustering with a terminating veto, is able to resolve nearby hard partons into separate jets, while it resembles common jet algorithms if the partons are well isolated. In addition to intrinsic jet properties, it introduces a dependence of the clustering history on two-jet properties, all formulated in terms of jet mass and mass ratios. It is this flexibility that outputs jets with variable effective radii, which leads to superior results compared to the fixed-radius variants.

While a χ^2-like measure could give a more accurate assignment of the jets to the various buckets, we gave an argument that the difference to our computationally inexpensive ansatz is not expected to be large. Another possible improvement is to require a certain number of b-tagged jets for each top and Higgs candidate. We did not include this option in our analysis because it would require matching between tagged jets and mass-jump jets, which have not yet been investigated by the experimental collaborations. Our results give a conservative estimate in this respect.

On top of the phenomenological study of vectorlike top pair production, we investigated the quality of reconstruction of the top quarks and Higgs bosons. Whereas the majority of tagging algorithms for boosted resonances assumes their isolation, we showed that our approach performs excellently in identifying the correct jet combinations even in this very busy and unclean environment. This study enters uncharted and often neglected territory when it comes to taggers. Yet the results are promising and we expect that jet clustering algorithms with a terminating veto will find their place in future studies of high-multiplicity processes.

5.A An analysis with adaptive fat jets

In this appendix we perform a numerical analysis of our benchmark process, $pp \to T\bar{T} \to t\bar{t}HH \to 10$ jets, based on fat jets. We argued in Sec. 5.3 that there are unavoidable problems with the definition of a

fat jet in this scenario, as the decay products of the boosted top quarks and Higgs bosons are not well separated. In particular the choice of radius for the fat jets has to be a compromise, and we speculated that the performance of tagging algorithms is thus impeded. In the following we describe a new, adaptive approach.

Version 2 of the `HEPTopTagger` includes a procedure that automatically adjusts the fat jet radius, labelled `optimalR mode` [85]. First we briefly review the algorithm. For decreasing fat jet radii from $R_{\text{max}} = 1.8$ down to $R_{\text{min}} = 0.5$ in steps of 0.1, the `HEPTopTagger` is applied and a top candidate constructed (the algorithm was reviewed in Sec. 3.6). The successive fat jets are obtained by reclustering the initial one with $R = 1.8$. The respective candidate's mass is labelled $m_{\text{rec}}^{(R)}$. It is expected that the series of values $m_{\text{rec}}^{(R)}$ observes a stable plateau until R is too small to capture all three subjets from the top decay. The smallest radius on the plateau is then defined as R_{opt} through the condition

$$\frac{m_{\text{rec}}^{(R_{\text{max}})} - m_{\text{rec}}^{(R)}}{m_{\text{rec}}^{(R_{\text{max}})}} > 0.2 \quad \Leftrightarrow \quad R < R_{\text{opt}}\,, \tag{5.27}$$

and the reconstructed momentum from this "optimal" fat jet forms the top candidate.

We anticipate that this adaptive top tagging algorithm may help in our scenario, as it can adjust the size of each fat jet individually. A similar Higgs tagger we define by porting the `optimalR mode` to the `BDRS Higgs tagger` (cf. Sec. 3.5). As the `BDRS` algorithm is agnostic of the Higgs boson mass and will yield very large masses for large fat jets due to splash-in, we modify the mass plateau condition in Eq. (5.27) as follows. While proceeding to smaller fat jet radii, the first candidate with a mass around the true Higgs mass,

$$m_{\text{rec}} \leq m_H + 20\,\text{GeV}\,, \tag{5.28}$$

determines the true starting point $R_{\text{max}'}$ for the `optimalR mode`. We then find the smallest fat jet on the mass plateau by Eq. 5.27 with R_{max} replaced by $R_{\text{max}'}$.

Our exemplary analysis proceeds as follows. Note that all cuts are very similar to those in our novel analysis of Sec. 5.4 to allow for easy comparison of the results.

1. Event preselection cuts:
   Scalar transverse momentum $H_T \geq 1400\,\text{GeV}$ and number of b-tagged jets $\#b \geq 4$.
   We refer to our novel analysis in Sec. 5.4.2 for details.

2. Build fat jets with $R = 1.8$ from the calorimeter information and apply the `HEPTopTagger` with `optimalR mode` starting from the leading fat jet. Once a top candidate with $p_\perp \geq 200\,\text{GeV}$ is reconstructed, remove its constituents from the calorimeter signal and repeat step 2. Only the two candidates with smallest $\Delta m = |m_{\text{rec}} - m_t|$ are finally kept (if there are less than two candidates, the event is discarded). Note that these candidates are not necessarily tagged: applying cuts after the identification of candidates reduces background shaping.[9] We do not modify the default cuts of the `HEPTopTagger`, in particular the allowed mass window for top candidates is $m_t \pm 25\,\text{GeV}$, the same as in Sec. 5.4.2.

[9] For the same reason, all tagging algorithms introduced in Sec. 3.6 apply cuts only after the candidate has been fixed. A similar philosophy is behind our original analysis described in Sec. 5.4 and the `HPTTopTagger` of Chapter 6.

3. Do step 2 for the `BDRS Higgs tagger` with `optimalR mode` as defined above. Now we do not require two or more Higgs candidates. The initial fat jets are formed from all calorimeter information but the constituents of the two top candidates removed. The allowed mass window for Higgs candidates is $m_H \pm 20\,\text{GeV}$ as in Sec. 5.4.2.

We define three signal regions analogously to our original analysis, recall Tab 5.1. The resulting event numbers and signal significances are shown in Tab. 5.4 for two different initial fat jet radii, $R_{\text{max}} = 1.8$ and 1.5. First we observe that signal region 1 is barely populated, i.e. the two top candidates identified by the procedure described above are very likely to be both tagged. This statement also holds for the $bb\bar{b}\bar{b}$ background. It is indicated that due to the high multiplicity and scalar transverse momentum, plentiful top-like substructure is induced. Comparing the two different setups, we observe that background fake rates are much larger when starting from a very large $R_{\text{max}} = 1.8$. The corresponding signal rates, on the other hand, do not gain similar efficiency. This hints that while the `optimalR mode` may help to correctly reduce the size of signal fat jets, splash-in effects and the resulting sheer subjet multiplicity severely impede background rejection. Even the sophisticated, self-adaptive fat jet algorithm considered here does not facilitate the problem of optimal jet radius and fat jet identification in a satisfactory manner. When the results are compared to those of our novel analysis in Tabs. 5.2 and 5.3, we see that the analysis based on fat jets described above cannot compete with the mass-jump based setup of Sec. 5.4.

Process		$T\bar{T}$					b.g.	$bb\bar{b}\bar{b}$	$t\bar{t}$	$t\bar{t}b\bar{b}$	$t\bar{t}H$
		800 GeV	900 GeV	1.0 TeV	1.1 TeV	1.2 TeV					
number of events for $100\,\text{fb}^{-1}$											
$H_T \geq 1.4\,\text{TeV}$		507	306	167	86.7	43.6	25600	4130	20600	772	52.9
$\#b \geq 4$		356	217	118	60.8	30.6	1730	990	506	218	16.4
R_{max} = 1.8	SR1	0.1	0.1	0.2	0.2	0.0	0.2	0.0	0.1	0.0	0.0
	SR2	56.9	31.1	15.9	8.6	4.7	42.2	4.8	22.2	13.5	1.8
	SR3	15.0	9.6	6.0	3.5	1.9	3.7	0.6	1.7	1.2	0.2
	$S/\sqrt{B}$	10.60	6.01	3.26	1.81	0.97					
R_{max} = 1.5	SR1	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0
	SR2	44.3	24.1	12.7	6.9	3.7	29.9	3.1	15.4	10.0	1.4
	SR3	9.7	7.4	4.5	2.6	1.4	2.1	0.3	0.9	0.7	0.1
	$S/\sqrt{B}$	9.56	5.57	3.05	1.69	0.90					

Table 5.4: Expected event numbers for a fat-jet-based analysis. Numbers are given for an integrated luminosity of $100\,\text{fb}^{-1}$ at the LHC with $\sqrt{s} = 14\,\text{TeV}$. Results for the signal are shown separately for different values of the vectorlike top mass ranging from 800 GeV to 1.2 TeV. All relevant background processes as well as their sum ("b.g.") are given in the right-hand columns. The three signal regions (SR) are defined in Tab. 5.1. The significance is defined as $S/\sqrt{B}$, taken from the number of signal events S and number of background events B summed over all three signal regions.

— All my life I've had one dream:
to achieve my many goals.

Homer Simpson

6

Top quarks and gauge bosons in the highly boosted regime

This chapter is based on work by the author [C, F].

Heavy resonances are an integral part of many extensions of the Standard Model. The discovery of such heavy resonances is a primary goal at the LHC and future hadron colliders. When a particle with TeV-scale mass decays into electroweak-scale objects, these objects are highly boosted and their decay products are then strongly collimated, possibly to an extent that they cannot be resolved in the calorimeters of the detectors any more. Also recall the introductory discussion in Sec. 3.1. We develop taggers for electroweak-scale resonances by combining the good energy resolution of the hadronic calorimeter with the superior spatial resolution of the tracking detector. Using track-based techniques we reconstruct heavy W' and Z' bosons. The taggers show a good momentum-independent performance up to very large boosts. Using the proposed techniques will allow experiments at the LHC and a future hadron collider to significantly extend its reach in searches for heavy resonances. Lastly, we apply a track-based analysis to constrain the branching ratio of the rare Higgs boson decay $H \to ZA \to l^+l^- + \text{jets}$.

6.1 Introduction

The scale of new physics for many anticipated extensions of the Standard Model has already been pushed beyond $\mathcal{O}(1)$ TeV (cf. Fig. 2.3). If heavy TeV-scale resonances decay into electroweak-scale particles, which in turn have large branching ratios into quarks, i.e. $X_{\text{TeV}} \to Y_{\text{EW}} \to \text{jets}$, these quarks are likely to

be collimated in the lab frame. More precisely, the extent to which the decay products of the electroweak-scale resonances are collimated depends on the ratio between the mass of the heavy new physics resonance and the electroweak scale. For central production one finds $p_{\perp,Y} \sim m_X/2$. As a result, for the decay products of Y their angular separation scales like $\Delta R_{\text{jets}} \sim 4m_Y/m_X$. As electroweak-scale resonances have generically large branching ratios into quarks, the reconstruction and detailed analysis of hadronic final states is at the core of the upcoming LHC program.

The topo-clusters from which jets are constructed in experiments by the ATLAS collaboration become the bottleneck of the analysis of highly boosted resonances. The minimal transverse size for a cluster of hadronic calorimeter cells is 0.3×0.3 in (η, ϕ), reached if all energy after noise subtraction is concentrated in one cell. To discriminate two jets the angular separation of their axes in the detector has to be at least $\Delta R \geq 0.2$. The CMS collaboration instead uses particle-flow objects. The charged hadrons, electrons and muons that enter these objects are reconstructed from tracks in the tracker, while photons and neutral hadrons are reconstructed from energy deposits in the electromagnetic and hadronic calorimeters, respectively. The spatial resolution of jets built from particle flow can be greatly improved with respect to calorimeter jets, i.e. the use of tracking detectors and the electromagnetic calorimeter's fine granularity allow the measurement of charged hadrons and photons inside a jet, which together constitute $\sim 85\%$ of the jet energy. Nevertheless the jet-energy resolution deteriorates quickly for jets with $R < 0.2$ [161]. Hence, the way CMS uses its particle-flow objects currently results in a lower limit on the spatial resolution of jets, similar to ATLAS.

For most Standard Model processes at the LHC this angular resolution is sufficient to separate the decay products of electroweak resonances. However, when scales are vastly separated and either $4m_Y/m_X \ll 0.2$ or in general $p_{\perp,X} \gg m_X$, the angular separation of the decay products can be too small to resolve them individually. While the overall energy deposit of highly boosted resonances can still be measured, the substructure, i.e. the energy sharing between the decay products, becomes oblivious. Hence the ability to discriminate between a decaying resonance and QCD jets using jet substructure observables quickly deteriorates. Obviously, at a possible future 100 TeV proton-proton collider where larger m_X are probed and the rate for electroweak resonances with $p_{\perp,X} \gg m_X$ is bigger, this issue cannot be ignored.

To recover reconstruction efficiency for highly boosted resonances and extend the multi-purpose experiments' sensitivity in searches for heavy resonances to larger masses, smaller input objects to jet algorithms have to be used. Following Ref. [25], we propose to use tracks instead of either topo-clusters or particle flow objects as input for resonance reconstruction methods[1], thereby trading an accurate measurement of the subjet's energy against an improved spatial resolution of the jet's substructure. We develop and refine dedicated reconstruction procedures for highly-boosted electroweak-scale resonances, e.g. W/Z bosons and top quarks, designed to exploit this trade-off between a precise measurement of the jet's energy and the improved spatial resolution of tracks. While we will focus on subjet-based techniques the same approach can be used for jet shape observables [163].

[1] In an earlier proposal [162] the same approach for the electromagnetic calorimetry was discussed for W tagging.

6.2 Tagging highly boosted top quarks

6.2.1 The original `HPTTopTagger`

The `HPTTopTagger` of Ref. [25] reconstructs the three-prong substructure of hadronically decaying top quarks from the charged tracks inside a given fat jet. These particle trajectories can be determined in a tracking detector to very high radial precision. More specifically, the ATLAS inner tracking detector achieves an angular resolution of $\Delta\eta \approx 10^{-3}$ and $\Delta\phi \approx 0.3\,\text{mrad}$ for charged particles with $p_\perp = 10\,\text{GeV}$ [164], while maintaining a reconstruction efficiency of $> 78\%$ for tracks of charged particles with $p_\perp \geq 500\,\text{MeV}$ [165]. Such accuracy cannot be achieved with the hadronic calorimeter where it is not possible to separately resolve small jets with $\Delta R_{j_1,j_2} < 0.2$.

Tagging algorithms that rely solely on information from calorimeter towers on the other hand – such as all those reviewed in Chapter 3 – are therefore not applicable any more if the top daughter jets are strongly collimated. The track-based `HPTTopTagger` ports elements of the `HEPTopTagger` algorithm (see review in Sec. 3.6) to the high-energy regime. However, to avoid combinatorial issues due to the high multiplicity of tracks and the introduction of artificial mass scales in background events, cuts are only applied on one three-subjet configuration inside the large-radius fat jet. Hence we do not search for a top-like structure in every possible subjet combination. The bulk of the top identification is then achieved by comparing ratios of invariant mass combinations of the three subjets. For example, the ratio m_{23}/m_{123} corresponds to m_W/m_t in most hadronic top decays, where m_{23} is the invariant mass of the sub-leading and sub-sub-leading subjet in transverse momentum and m_{123} is the invariant mass of the top candidate.

Due to imperfect knowledge of all energy flow in the tracking detector (only charged particles are reconstructed), the track momenta are scaled according to the inverse of the energy fraction carried by charged tracks [25], [162]

$$\alpha_j \equiv \frac{E_{\text{jet}}}{E_{\text{tracks}}}\,. \tag{6.1}$$

Because the energy of the (hadronic) fat jet can be calibrated to good precision in the experiment [166], [167], the sensitivity to fluctuations is hence ameliorated. In essence, the tracker's spatial and the calorimeter's improved energy resolution are combined to perform a local recalibration of the track-based fat jet.

For completeness, we echo the complete algorithm as presented in Ref. [25]:

1. Define a jet j using the C/A algorithm with $R = 0.8$ from calorimeter clusters.
2. Take the tracks with $p_\perp > 500\,\text{MeV}$ that are associated with j and recombine them to a track-based jet j_c.[2]
3. Calculate α_j of Eq. (6.1) using j and j_c.
4. Apply the mass-drop procedure introduced in Refs. [24], [168]: undo the last clustering of the track-based jet j_c into two subjets j_{c1}, j_{c2} with $m_{j_{c1}} > m_{j_{c2}}$. We require $m_{j_{c1}} < 0.8\,m_{j_c}$ to keep j_{c1} and j_{c2}. If this condition does not hold we keep only j_{c1}. Each subjet j_{ci} we further decompose unless $m_{j_{ci}} < 20\,\text{GeV}$. The remaining subjets we add to the list of relevant substructures.

[2] We use ghost-association, cf. footnote 3 in Sec. 5.3.

5. If we find fewer than two remaining subjets we consider the tag to have failed. Else, we take the constituents of all subjets surviving the mass-drop procedure and multiply their momenta by α_j each.

6. We take all the rescaled constituents and filter (cf. Sec. 3.3) them with resolution $R_{\text{filt}} = \max(\,0.05, \min(\Delta R_{ij}/2)\,)$, in which i and j run over all remaining subjets after the mass-drop procedure. We recombine all constituents of the four hardest filtered subjets and require the resulting invariant mass to be in a mass window around the top quark mass. We call this object our top candidate.

7. Again we follow the `HEPTopTagger` and construct exactly three $p_\perp$-ordered subjets j_1, j_2, j_3 from the top candidate's constituents. If the masses (m_{12}, m_{13}, m_{23}) satisfy the so-called A-cut of Eq. (3.10), we consider the top tag to be successful.

6.2.2 The new default of the `HPTTopTagger`

The `HPTTopTagger` described above is essentially a track-based version of the `HEPTopTagger`, with reasonable adjustments to account for very large transverse momenta but no major revisions. In the following, we describe key observations that lead to an improved top tagger in the highly boosted regime. All analyses in this chapter are based on this new default.

First we consider energy calibration of the charged tracks in Eq. (6.1). In fact, calibrations of Cambridge-Aachen (C/A) subjets are available for radius parameters as small as 0.2 [167] which makes it beneficial to rescale the charged tracks more locally. For boosted tops with transverse momentum at the TeV scale, typically two or all three top subjets can be resolved in this way and the fluctuations are expected to be reduced separately. In Fig. 6.1 we show the average mutual separation between top decay products. Thus, among other improvements, the new default of the `HPTTopTagger` applies the approach of Eq. (6.1) to subjets individually instead of the whole fat jet.

We find that the Monte Carlo truth partons from the decay $t \to bW^+ \to bjj$ are separated by a characteristic R distance of $\sim 200\,\text{GeV}/p_{\perp,t}$ and that the energy carried by charged tracks around these partons is very well localized with a much smaller radius, see Fig. 6.2. Given these observations, we are led to abandon the mass-drop unclustering procedure, which was inherited from the `HEPTopTagger`, in favour of conventional (anti-k_T) subjets with radius parameter $R = 100\,\text{GeV}/p_{\perp,j_c}$. We label the three subjets leading in transverse momentum $\tilde{j}_1$, $\tilde{j}_2$ and $\tilde{j}_3$. This procedure renders an additional filtering stage redundant.

The inclusion of the leading gluon emissions is important to capture all top quark decay products, in particular when $p_{\perp,t} \gg m_t$. In Ref. [25] this was not done explicitly, which resulted in relatively low reconstructed masses. Additional soft emissions can, however, easily be captured if we allow our subjets to overlap, i.e. we recluster the fat jet j_c with $R = 0.8 \cdot \min \Delta R(\tilde{j}_i, \tilde{j}_k)$ where $i, k \in \{1, 2, 3 \mid i \neq k\}$. As the new radius is smaller than the subjet separation, it is guaranteed that they are still resolved separately. The new $p_\perp$-leading subjets j_1, j_2 and j_3 form the top candidate.

Typically tagging efficiencies will be larger for the new `HPTTopTagger` which is a desired feature, especially when the luminosity of the experiment is limited. To achieve good rejection rates of background

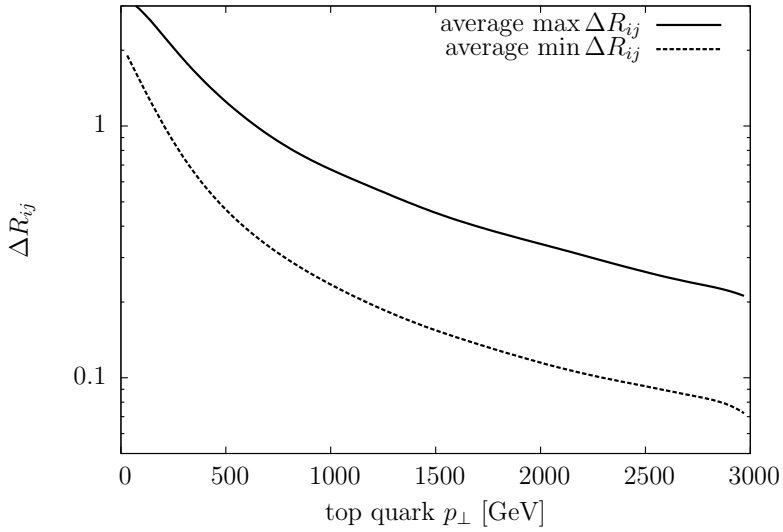


Figure 6.1: Average smallest and largest mutual separation in $\Delta R_{ij} \equiv \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2}$ between the partons from the hadronic top decay $t \to bW^+ \to bjj$.

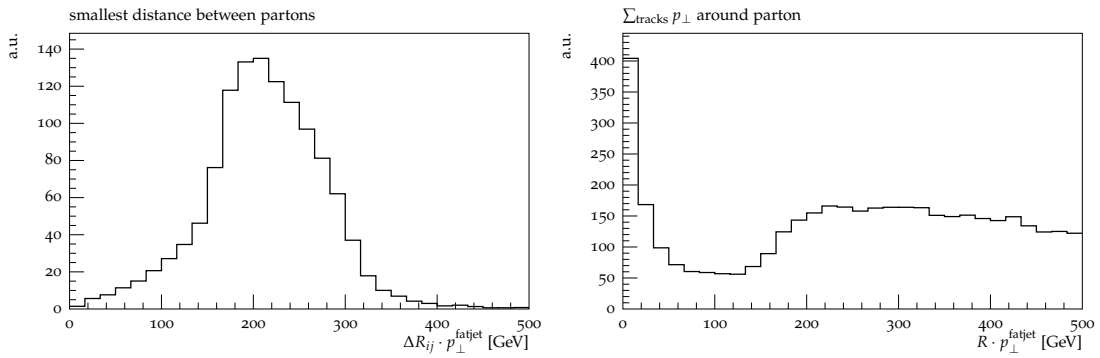


Figure 6.2: Smallest separation between two MC truth partons in a top decay, scaled with the fat jet transverse momentum (left) and distribution of transverse momentum carried by charged tracks around the truth partons (right).

events, in addition to the so-called A-cut of mass ratios [cf. Eq. (3.10)], we apply a cut based on the geometric size of the top candidate. The scaled spread of the reconstructed subjets has to satisfy

$$\max \Delta R(j_i, j_k) \cdot p_{\perp,\text{top candidate}} \in [\,250\,\text{GeV}, 750\,\text{GeV}\,] \tag{6.2}$$

where $i, k \in \{1, 2, 3 \mid i \neq k\}$, in order to reject fake candidates where the big top-like mass is generated through soft large-angle radiation.

We define the new default of the `HPTTopTagger` algorithm by the following procedure:

1. Define a fat jet j using the C/A algorithm with $R = 0.8$ and $p_\perp \geq 800\,\text{GeV}$.[3]
2. Discard all tracks that are not associated with j or that have $p_\perp < 500\,\text{MeV}$.[4]
3. Scale the remaining track momenta as follows: Recluster j with the anti-k_T algorithm employing a small radius $R = 0.2$, and calculate $\alpha_j \equiv E_\text{jet}/E_\text{tracks}$ for each subjet using its respective associated tracks. We multiply the momenta of the tracks by α_j and recombine the scaled tracks to a track-based jet j_c.
4. Recluster j_c using the anti-k_T algorithm with $R = 100\,\text{GeV}/p_{\perp,j_c}$. If there are fewer than three subjets we consider the tag to have failed.
5. We then calculate the smallest pairwise distance between the three leading subjets, $r_\text{min} \equiv \min \Delta R_{ij}$ and recluster j_c with a new radius $R = 0.8\, r_\text{min}$. If the new three leading subjets result in an invariant mass around the top quark mass, $m_\text{candidate} \in m_t \pm 25\,\text{GeV}$, they form our top candidate.
6. We follow the `HEPTopTagger` and apply the so-called A-cut of Eq. (3.10) on the pairwise invariant masses (m_{12}, m_{13}, m_{23}). If in addition the top candidate satisfies $p_\perp \cdot \max \Delta R_{ij} \in [\,250\,\text{GeV}, 750\,\text{GeV}\,]$, we consider the top tag to be successful.

6.2.3 Performance

We investigate the tagging efficiency and the reconstructed mass of the `HPTTopTagger`, as described in Sec. 6.2.2. Throughout this section, we generate top-initiated jets from the production of a heavy boson Z', $pp \to Z' \to t\bar{t} \to$ jets, for masses in the range $m_{Z'} = 3 \cdots 6\,\text{TeV}$. Unless otherwise stated, the sample with mass $m_{Z'} = 6\,\text{TeV}$ is used to generate the plots. For the background we consider QCD jets from dijet production. Unless stated otherwise, we generate events with Pythia 8.186 at centre-of-mass energy $\sqrt{s} = 14\,\text{TeV}$. Fat jets are clustered using FastJet 3.0.6 from stable particles with pseudorapidity $|\eta| < 4.9$. We apply the C/A jet algorithm with radius parameter $R = 0.8$ and $p_\perp \geq 800$ GeV. To assess the signal efficiency we match the fat jet to MC truth top quarks by requiring $\Delta R(j, \text{top}) < 0.6$ and select jets with $|\eta_j| < 2.5$.

A priori, constructing observables from tracks can result in a marked sensitivity on detector resolution and efficiencies. To evaluate the impact of these effects on the performance of the algorithm we use the

[3] Note that the choice of fat jet clustering algorithm is less vital compared to previous approaches because we do not apply an unclustering step. The same holds for our W and Z boson taggers developed in Sec. 6.3.1.

[4] We use ghost-association, cf. footnote 3 in Sec. 5.3.

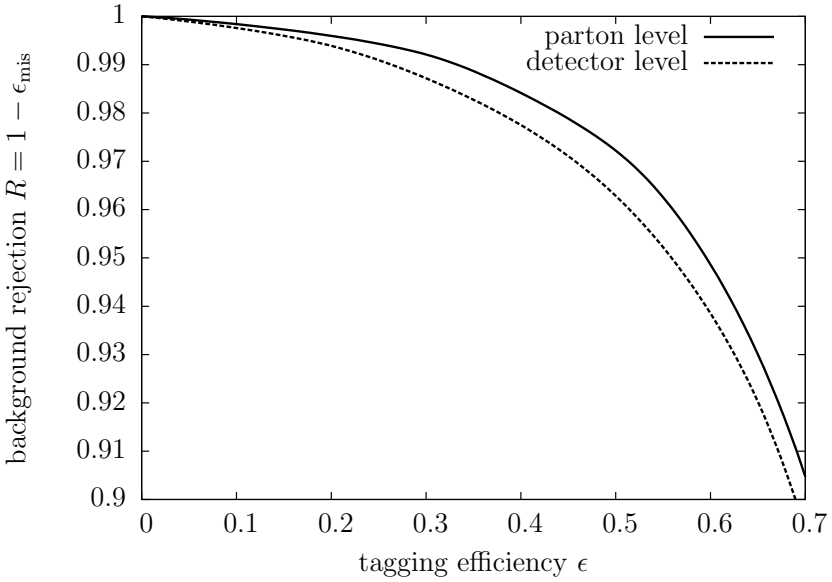


Figure 6.3: ROC curves for scan over m_{23}/m_{123}, $p_\perp \cdot \max \Delta R_{ij}$, m_t, m_W, Δm_t, Δm_W in various combinations and pre-scan cuts. For details on signal and background generation see the text.

DELPHES 3.1.2 fast dectector simulation. Throughout parameters are chosen in accordance with the default ATLAS parameters provided by DELPHES.

Fig. 6.3 shows the receiver-operater characteristic (ROC) curve for our algorithm. By tuning the free parameters and cuts, background rejection $R = 1 - \epsilon_{\text{mis}}$ can be maximized for any working point with given signal tagging efficiency ϵ. We use TMVA [169] to find optimum cuts on m_{23}/m_{123}, $p_\perp \cdot \max \Delta R_{ij}$, as well as the allowed mass windows around m_t and m_W. TMVA assesses the cut parameters by their respective discrimination power and applies cuts in this order. To further improve performance, we run the algorithm for different combinations of fixed and free cut parameters and combine the resulting curves such that at each working point, the setup with the largest background rejection is selected. To obtain a dropping $p_\perp$ distribution as expected in a real analysis, we take background QCD jets from dijet production with $\hat{p}_\perp \geq 700\,\text{GeV}$. $\hat{p}_\perp$ is the transverse momentum in the rest frame of the hard process at generator level.

In Fig. 6.4 we show the $p_\perp$-dependent tagging efficiencies of the `HPTTopTagger`. Here we obtain background QCD jets from dijet production with binned generator-level $\hat{p}_\perp$ in the range $[700, 2500]\,\text{GeV}$ and $\hat{p}_\perp \geq 2500\,\text{GeV}$ to achieve good statistics in all bins. Over the whole studied $p_\perp$ range we find a flat tagging efficiency and fake rate. The outlined modifications for the new default of the `HPTTopTagger` improve the tagging efficiency for fat jets with $p_{\perp,t} \sim 1 - 2$ TeV while maintaining a similar fake rate. For particle-level final states with $p_{\perp,\text{fatjet}} \sim 1$ TeV the `HPTTopTagger` has a signal efficiency of roughly 35% which slightly decreases to 30% at $p_{\perp,\text{fatjet}} = 3$ TeV. Including detector effects results in a flat shift to lower values by 5% over the entire $p_{\perp,\text{fatjet}}$ range.

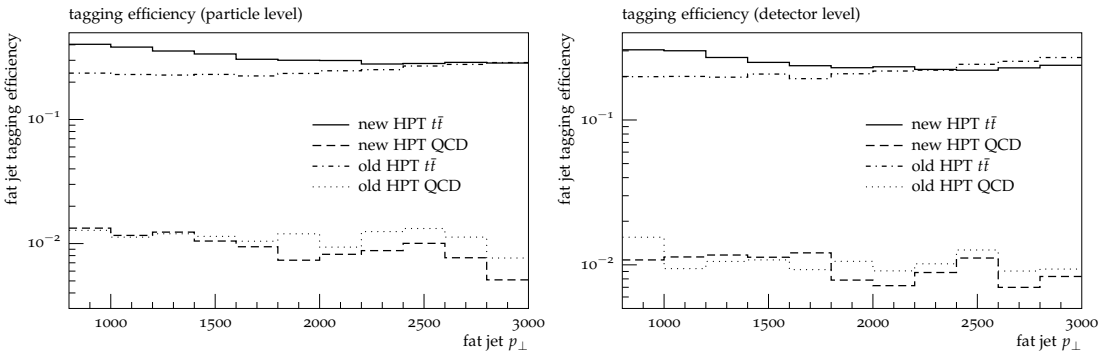


Figure 6.4: Tagging efficiency of the original `HPTTopTagger` and the new default at particle level (l.h.s.) and detector level (DELPHES) (r.h.s.).

Fig. 6.5 shows the reconstructed top mass after applying the `HPTTopTagger` in two different transverse momentum windows, $p_{\perp,\text{fatjet}} \in [1000, 1500]\,\text{GeV}$ (l.h.s.) and $[2000, 2500]\,\text{GeV}$ (r.h.s.). In the former, we generate background events with $\hat{p}_\perp \geq 700\,\text{GeV}$ and in the latter with $\hat{p}_\perp \geq 1800\,\text{GeV}$, to obtain a dropping $p_\perp$ distribution as expected from the SM process. The top mass distribution from the signal sample shows a clear peak at the correct top quark mass. While the detector simulation does not affect the position and width of the peak significantly, it is slightly sharper and more pronounced for $p_{\perp,\text{fatjet}} \in [1000, 1500]$ GeV compared to the window [2000, 2500] GeV.

As track-based observables are not infrared safe [170]–[173] non-perturbative contributions have to be taken into account to obtain well-defined results. For practical purposes, in event generators, hadronization models including fragmentation functions are used, which in turn follow perturbative evolution equations to obtain an extrapolated result at energies of interest from LEP data [174]–[177]. Thus, track-based observables are subject to the parametrisation of non-perturbative physics.

To estimate the impact different hadronisation models can have on the performance of the `HPTTopTagger` we compare the tagging efficiency and mass reconstruction for events generated with either PYTHIA 8.186 or HERWIG++ 2.7.1 [178], see Fig. 6.6. We find that both event generators result in very similar signal efficiencies, i.e. differences of $\mathcal{O}(10)\%$ at most. For the backgrounds the differences are slightly larger but still within generic uncertainties of event generators.

6.2.4 Resonance search with highly boosted top quarks

The discovery of heavy resonances is a prime goal at the LHC and a possible future high-energy proton-proton collider. In many extensions of the Standard Model heavy resonances are predicted (cf. Sec. 2.3), leading to highly boosted top quarks. While for top quarks with intermediate boost focusing on (semi)leptonic top decays can lead to stronger exclusion limits [179], at very high transverse momentum standard lepton-isolation requirements fail and the improved background rejection in leptonic final states ceases to compensate for the smaller branching ratio. Thus, we will again focus on hadronic top decays only, as we do throughout this thesis.

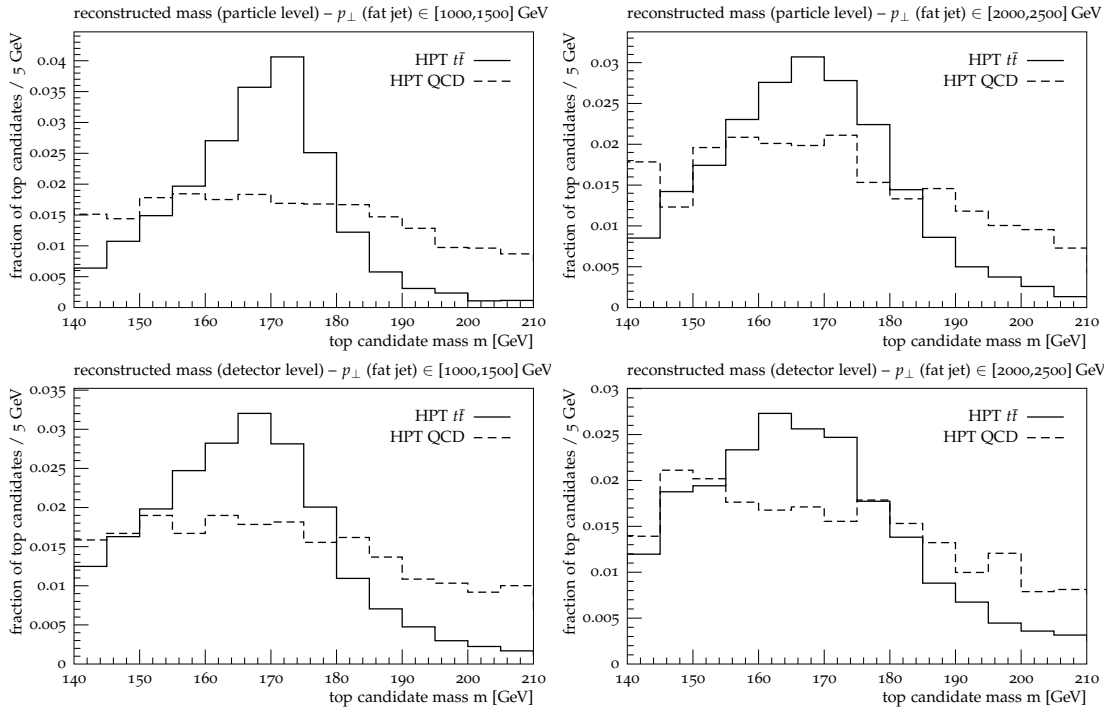


Figure 6.5: Reconstructed top quark mass with the new `HPTTopTagger` at particle level (upper row) and detector level (DELPHES) (lower row) for different transverse momentum ranges $p_{\perp,\text{fatjet}}$.

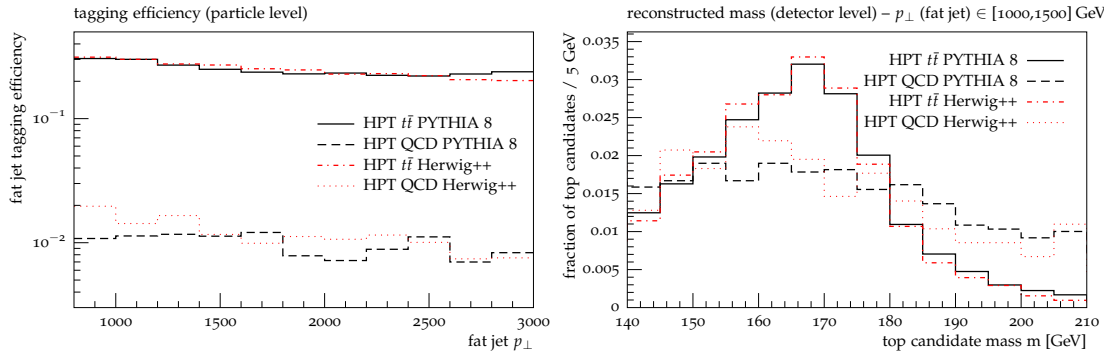


Figure 6.6: Comparison of results for the new `HPTTopTagger` from samples generated with PYTHIA (black curves) and HERWIG++ (red curves). Tagging efficiencies and reconstructed mass are shown at detector level for both top-initiated and QCD jets.

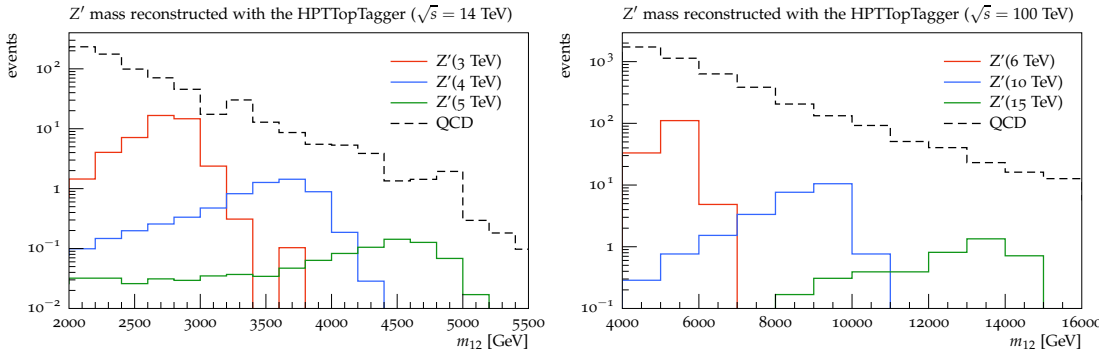


Figure 6.7: Invariant mass of the two reconstructed top quarks after top tag m_{12}. The left figure corresponds to benchmark scenarios for $\sqrt{s} = 14$ TeV and the right figure to $\sqrt{s} = 100$ TeV with 300 fb^{-1} integrated luminosity each.

Results for LHC

Fig. 6.7 shows expected event rates at the LHC with 14 TeV centre-of-mass energy and an integrated luminosity of 300 fb^{-1}. We generate three signal benchmark scenarios with PYTHIA, choosing $m_{Z'} = (3, 4, 5)$ TeV and universal default couplings to fermions. The resonance widths are $\Gamma_{Z'} = (95, 128, 160)$ GeV respectively, and the branching ratio BR($Z' \to t\bar{t}$) gives 10.5–10.6% in the considered mass regime. Resulting production cross-sections are $\sigma_{Z'} = (3.45, 0.501, 0.00962)$ fb respectively, which only include events where the heavy Z' decays into a pair of top quarks, and each top decays hadronically. We generate QCD dijet background events with $\hat{p}_\perp \geq 700$ GeV and find a total cross-section of 111.6 pb.

To reconstruct the Z' resonances we require at least two fat jets clustered using the C/A jet algorithm with $R = 0.8$. Input to the jet finder are all final state particles with $|\eta| < 4.9$, processed through DELPHES. We further impose $p_{\perp,\text{fatjet}} \geq 800$ GeV.

If both fat jets are tagged we reconstruct the resonance by summing the reconstructed top quarks' four-momenta, $m_{12}^2 = (p_{t_1} + p_{t_2})^2$. The results are plotted in Fig. 6.7. While the resonance peak is clearly visible, it is shifted to values below the true resonance mass. In addition, for all resonance masses we find a long tail towards smaller m_{12}. The shift of the mass peak and the low-mass tail are more pronounced for heavier resonances. This is a known effect when reconstructing very heavy resonances. Top taggers aim to reconstruct top quarks that are close to being on-shell. However, top quarks produced in decays of heavy resonances have a large probability to radiate gluons before decaying. As these gluons are emitted in a wide angle around the top quark this radiation is lost when recombining the four-momenta of the reconstructed tops and their invariant mass is shifted towards smaller values.[5]

We evaluate the required integrated luminosity to exclude the three benchmark resonances using a simple cut and count method. Based on the invariant mass distribution of the two reconstructed top

[5] A possible solution to this problem is discussed in Ref. [85]. For the same signal process with a moderately heavy Z' of mass $m_{Z'} = 1.5$ TeV it was found that using the filtered fat jets (not only the tagged substructure) can restore the Z' mass peak. However, background events are also shifted to larger masses, resulting in no improvement in Z' extraction from this observable alone.

Resonance	m_{12} window (TeV)	S/B	$S/\sqrt{B}$	σ (fb)	σ for $S/\sqrt{B} = 2$
$m'_Z = 3$ TeV	2.6-3.0	0.27	2.90	3.45	2.38
$m'_Z = 4$ TeV	3.4-4.0	0.13	0.69	0.501	1.45
$m'_Z = 5$ TeV	4.4-5.0	0.07	0.16	0.00962	0.12

Table 6.1: Results for search for $Z' \to t\bar{t}$ at the LHC14 in three benchmark scenarios. The last column shows the required production cross-section to achieve $S/\sqrt{B} = 2$ with 300 fb^{-1} integrated luminosity. Note that σ includes the branching fractions BR($Z' \to t\bar{t}$) and BR($t/\bar{t} \to$ jets). All numbers are based on the results provided in Fig. 6.7.

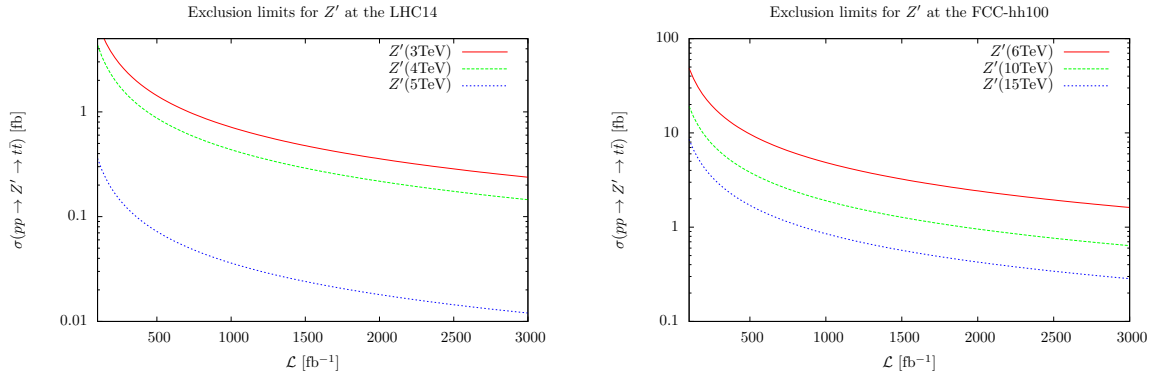


Figure 6.8: Exclusion limits at 90% CL for $Z' \to t\bar{t}$ with the new `HPTTopTagger` depending on production cross-section and integrated luminosity. The left figure corresponds to benchmark scenarios for $\sqrt{s} = 14$ TeV and the right figure to $\sqrt{s} = 100$ TeV.

quarks m_{12} we choose the two bins with the best S/B ratio and require $S/\sqrt{B} \simeq 2$.

For our benchmark scenarios, with only 300 fb^{-1} of data, a heavy $m_{Z'} = 3\,\text{TeV}$ can be excluded at 90% confidence level (CL).

Based on Tab. 6.1, in Fig. 6.8 we show the necessary integrated luminosity for exclusion at 90% CL, depending on the resonance's production cross-section.

Results for FCC-hh 100 TeV

We generate events as described above but with a centre-of-mass energy of $\sqrt{s} = 100\,\text{TeV}$, i.e. we consider the hadronic final state of top quark pair production via a heavy gauge boson Z' over a QCD dijet background. Note that detector specifications are identical to the LHC analysis for concreteness. We employ the same analysis to reconstruct the heavy Z' from two tagged tops, but now select fat jets with $p_{\perp,\text{fatjet}} \geq 2000$ GeV.

Again, we generate three signal benchmark scenarios with Pythia as detailed above, choosing bench-

Resonance	m_{12} window (TeV)	S/B	$S/\sqrt{B}$	σ(fb)	σ for $S/\sqrt{B}=2$
$m'_Z = 6$ TeV	5.0-6.0	0.098	3.29	26.9	16.39
$m'_Z = 10$ TeV	8.0-10.0	0.054	0.99	3.18	6.45
$m'_Z = 15$ TeV	13.0-15.0	0.052	0.33	0.466	2.84

Table 6.2: Results for search for $Z' \to t\bar{t}$ at the FCC-hh100 in three benchmark scenarios. The last column shows the required production cross-section to achieve $S/\sqrt{B} = 2$ with 300 fb^{-1} integrated luminosity. Note that σ includes the branching fractions BR($Z' \to t\bar{t}$) and BR($t/\bar{t} \to$ jets). All numbers are based on the results provided in Fig. 6.7.

mark masses $m_{Z'} = (6, 10, 15)$ TeV. Production cross-sections turn out $\sigma_{Z'} = (26.9, 0.986, 0.328)$ fb and resonance widths $\Gamma_{Z'} = (192, 322, 485)$ GeV respectively. We generate QCD dijet background events with $\hat{p}_\perp \geq 2000\,\text{GeV}$ and find a total cross-section of 149.3 pb.

Expected event rates at $300\,\text{fb}^{-1}$ are shown in Fig. 6.7 (r.h.s.) for all three benchmark masses and the QCD background. The mass bins that give the largest significance are listed in Tab. 6.2. Based on this table, we show 90% CL exclusion limits as a function of the production cross-section and integrated luminosity in Fig. 6.8 (r.h.s.).

6.3 Tagging highly boosted gauge bosons

6.3.1 The `HPTWTagger` and `HPTZTagger` algorithms

Not only boosted top quarks are of interest for searches of new physics at the LHC and future colliders. Many extensions of the Standard Model predict heavy resonances decaying to W or Z bosons (also see Sec. 2.3). In addition, SM cross-sections for the emission of electroweak gauge bosons off quarks are strongly enhanced in energetic final states [180]–[182]. Thus, in this section we outline a proposal for a track-based tagger for highly boosted W and Z bosons.

In the hadronic decay of boosted electroweak gauge bosons, $W^\pm \to q\bar{q}'$ or $Z \to q\bar{q}$, the two daughter partons exhibit a typical separation $\Delta R \sim 2m_{W/Z}/p_{\perp,W/Z}$. A successful tagging algorithms has to resolve these subjets separately, where the finite resolution of the hadronic calorimeter becomes the bottleneck for large transverse momenta $p_\perp \gtrsim 500\,\text{GeV}$. Following the arguments above, it is straightforward to adapt the `HPTTopTagger` to W and Z boson decays. Due to the two-prong nature of the process, we replace the cut on mass ratios ("A-cut") by a cut on the momentum fraction of the leading subjet,

$$f_{p_\perp} \equiv \frac{p_{\perp,j_1}}{p_{\perp,W/Z\text{ candidate}}} = \frac{p_{\perp,j_1}}{p_{\perp,j_1+j_2}}\,. \tag{6.3}$$

Fake candidates from QCD jets can generate a W- or Z-like mass via soft emissions. In this case $f_{p_\perp}$ is thus dominated by values close to 1.

We propose the following algorithm to reconstruct highly boosted W and Z bosons:

1. Define a fat jet j using the C/A algorithm with $R = 0.5$ and $p_\perp \geq 500\,\text{GeV}$.

2. Discard all tracks that are not associated with j or that have $p_\perp < 500\,\text{MeV}$.[6]

3. Scale the remaining track momenta as follows: Recluster j with the anti-k_T algorithm employing a small radius $R = 0.2$, and calculate $\alpha_j \equiv E_{\text{jet}}/E_{\text{tracks}}$ for each subjet using its respective associated tracks. The momenta of those tracks are multiplied by α_j. Combine the scaled tracks associated with the fat jet to a track-based jet j_c.

4. Recluster j_c using the anti-k_T algorithm with $R = 80\,\text{GeV}/p_{\perp,j_c}$. If there are fewer than two subjets we consider the tag to have failed.

5. Calculate the distance between the two leading subjets, $r \equiv \Delta R_{ij}$ and recluster j_c with a new radius $R = 0.8\,r$. If the combination of the new two leading subjets gives an invariant mass around the electroweak boson mass, $m_{\text{candidate}} \in m_{W/Z} \pm 10\,\text{GeV}$, they form our boson candidate.

6. If the momentum fraction of the two candidate subjets as defined in Eq. (6.3) satisfies $f_{p_\perp} \leq 0.85$ (for W candidates) or $f_{p_\perp} \leq 0.80$ (for Z candidates), we consider the electroweak boson tag successful.

If higher tagging efficiencies are required, we suggest to modify steps 5 and 6 and allow a larger candidate mass window, $m_{\text{candidate}} \in m_{W/Z} \pm 15\,\text{GeV}$, together with a looser cut on the subjet momentum fraction, $f_{p_\perp} \leq 0.85$. This setup is denoted "working point 2" (w.p. 2) below.

We are not concerned with discriminating between W and Z candidates. While the mass peaks are fairly well separated (cf. Figures 6.10 and 6.13), one may also consider additional distinguishing features such as (sub)jet charge [183].

6.3.2 Performance

We investigate the tagging efficiency and the reconstructed mass of the `HPTWTagger` as well as the `HPTZTagger`, as described in Sec. 6.3.1.

The relevant signal fat jets for both taggers are taken from the production of a hypothetical charged heavy boson at the LHC, decaying into a W and a Z boson,

$$pp \to W'^{\pm} \to W^{\pm} Z\,. \tag{6.4}$$

We generate events with masses in the range $m_{W'} = 3 \cdots 6\,\text{TeV}$. We consider two semileptonic decay patterns, $pp \to W'^{\pm} \to W^{\pm} Z \to (jj)(l^+l^-)$ to assess the `HPTWTagger`, and $pp \to W'^{\pm} \to W^{\pm} Z \to (l^{\pm}\nu)(jj)$ for the `HPTZTagger`, respectively. Background events with a generator-level cut $\hat{p}_\perp = 400\,\text{GeV} \cdots 2500\,\text{GeV}$ are generated from $pp \to Zj \to (l^+l^-)j$ for the `HPTWTagger`. The background process to assess the `HPTZTagger` is $pp \to W^{\pm} j \to (l^{\pm}\nu)j$ with the same generator-level cuts. We generate all events with Pythia 8 at centre-of-mass energy $\sqrt{s} = 14\,\text{TeV}$. Fat jets are clustered from stable particles with pseudorapidity $|\eta| < 4.9$ using the C/A jet algorithm with radius parameter $R = 0.5$ and $p_\perp \geq 500$ GeV. To assess the signal efficiency we match the fat jet to MC truth W / Z bosons by requiring $\Delta R(j, \text{W/Z}) < 0.4$ and select jets with $|\eta_j| < 2.5$. Additionally, we require the fat jet to be isolated from the MC truth leptons from the other gauge boson. For a leptonically decaying Z

[6] We use ghost-association, cf. footnote 3 in Sec. 5.3.

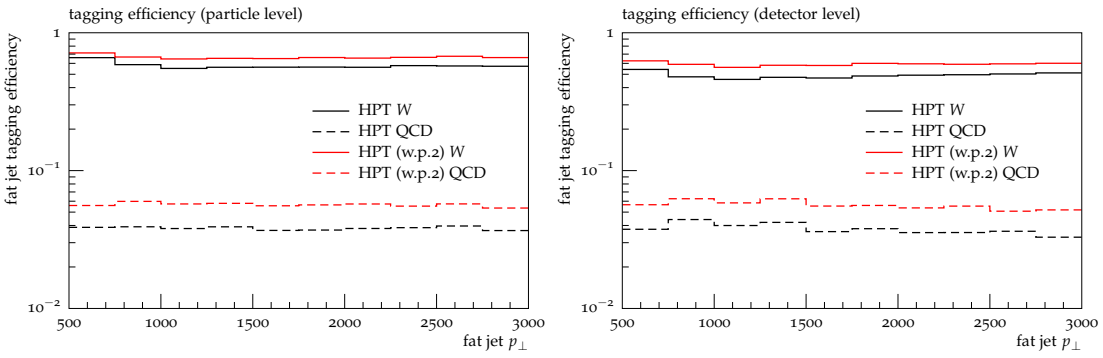


Figure 6.9: Tagging efficiency of the `HPTWTagger` at particle level (l.h.s.) and detector level (DELPHES) (r.h.s.) for boosted W bosons (solid lines). Standard Model QCD background mistag rates are given by the dashed lines. Results are shown for the standard setup (black) and the higher-efficiency working point 2 (red).

boson, the condition is $\Delta R(j, \text{lepton}) > 0.6$. The isolation criterion for a leptonic W boson decay reads $\Delta R(j, \text{lepton/neutrino}) > 0.6$.

Fig. 6.9 shows the efficiencies for tagging highly boosted W bosons with the `HPTWTagger` and background mistag rates for both working points suggested in Sec. 6.3.1. Results are presented at particle level (left) and detector level after running DELPHES (right). We use all event samples specified above to generate the plots in order to achieve good statistics over the whole $p_\perp$ range.

We show the reconstructed W boson mass in Fig. 6.10 in two different $p_\perp$ bins, also both at particle and detector level. In the $p_\perp \in [1000, 1500]\,\text{GeV}$ bin (left), we only consider the $m_{W'} = 3\,\text{TeV}$ sample for signal events. In order to obtain a dropping $p_\perp$ distribution as expected from the SM process, we generate background events with fixed generator-level cut $\hat{p}_\perp \geq 1000\,\text{GeV}$. At higher transverse momenta in the $p_\perp \in [2000, 2500]\,\text{GeV}$ bin (right), we simulate signal events with $m_{W'} = 5\,\text{TeV}$ and for the background we impose a generator-level $\hat{p}_\perp \geq 2000\,\text{GeV}$.

In Fig. 6.11 we compare the tagging efficiency of our `HPTWTagger` to a `BDRS`-like algorithm as defined in Sec. 3.5 with default parameters. In accordance with out algorithm, we consider the tag successful if the invariant mass lies within $m_W \pm 10\,\text{GeV}$. The performance of the `BDRS` algorithm as implemented according to Eqs. (3.8) and (3.9) quickly deteriorates for increasing jet $p_\perp$, see Fig. 6.11. For $p_{\perp,j} > 1500$ GeV it is more likely to obtain a W-tag with a QCD jet than with a W decay. While optimising the parameters entering Eq. (3.8) might help to recover efficiency, the minimal angular separation of calibrated filtered subjets $\Delta R_{j1,j2} > 0.2$ (for filtering see Sec. 3.3) eventually limits the applicability of this tagger for highly boosted resonances.

We show the boosted Z boson tagging efficiencies of the `HPTZTagger` in Fig. 6.12 for both working points and both at particle level (left) and detector level (right). Again, tagging efficiency and background rejection are stable over a large range of transverse momentum from $p_\perp = 500\,\text{GeV}$ up to 3 TeV. Throughout the remainder of this section, the cuts imposed at generator level, i.e. signal $m_{W'}$ and $\hat{p}_\perp$ for the background, are the same as in the corresponding plots of the `HPTWTagger`.

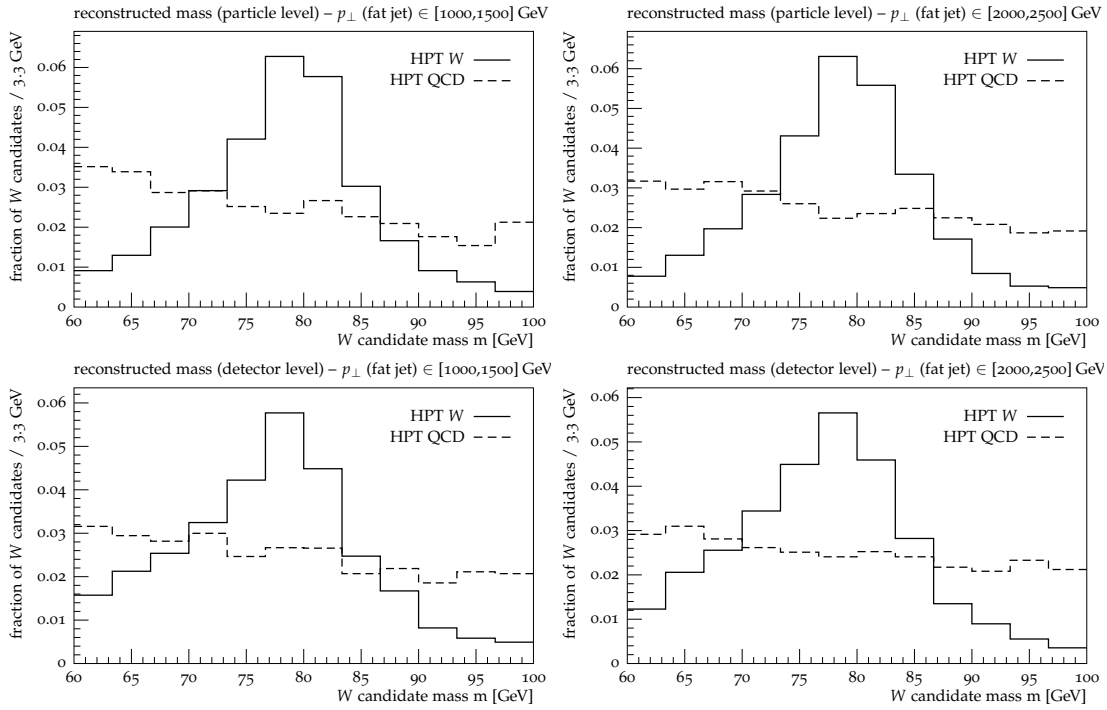


Figure 6.10: Reconstructed W boson mass with the `HPTWTagger` at particle level (upper row) and detector level (DELPHES) (lower row) for different $p_{\perp,\text{fatjet}}$.

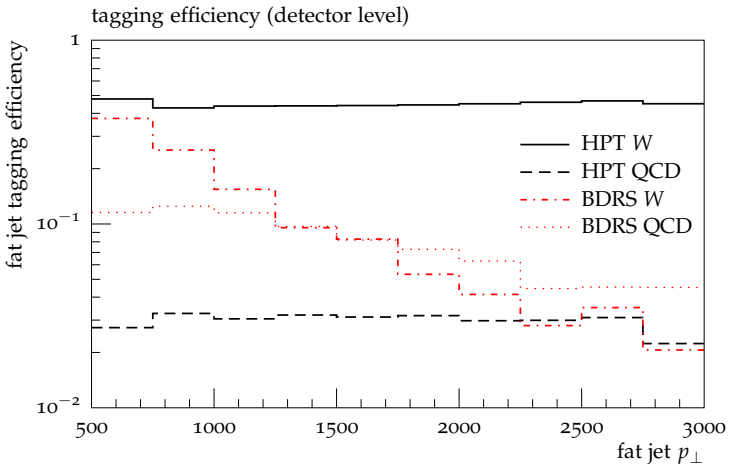


Figure 6.11: Comparison of the tagging efficiencies of the `HPTWTagger` (black lines) and a `BDRS`-like tagger (red lines) at detector level (DELPHES) for boosted W bosons. Standard Model QCD background mistag rates are also given.

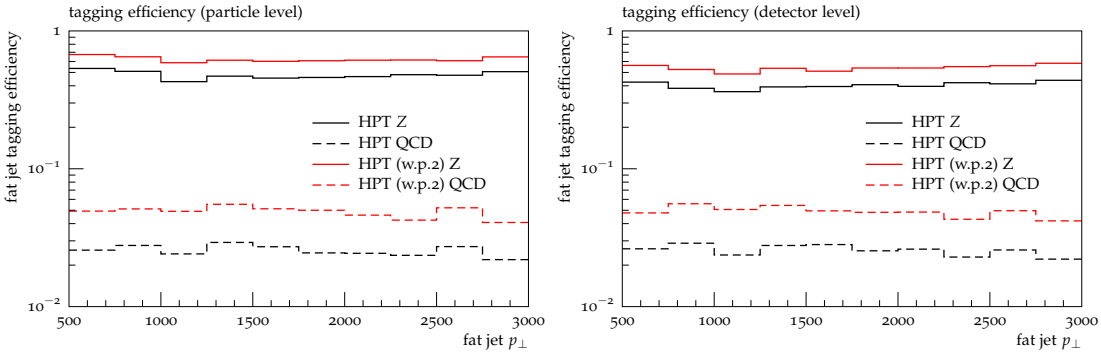


Figure 6.12: Tagging efficiency of the `HPTZTagger` at particle level (l.h.s.) and detector level (DELPHES) (r.h.s.) for boosted Z bosons (solid lines). Standard Model QCD background mistag rates are given by the dashed lines. Results are shown for the standard setup (black) and the higher-efficiency working point 2 (red).

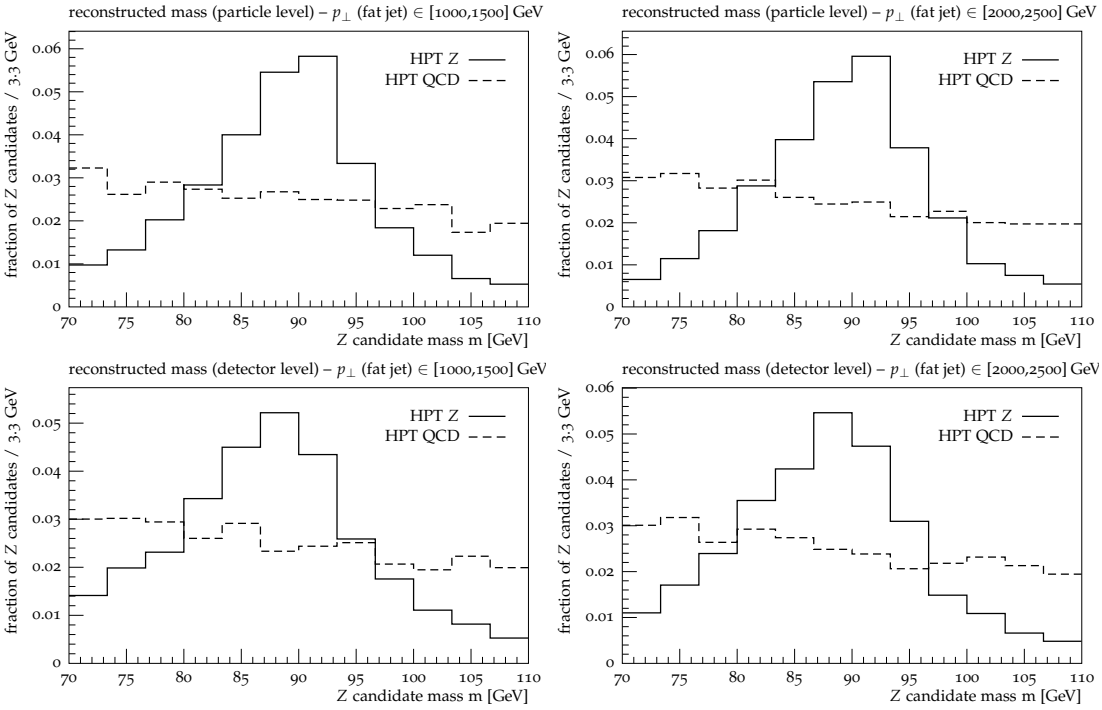


Figure 6.13: Reconstructed Z boson mass with the `HPTZTagger` at particle level (upper row) and detector level (DELPHES) (lower row) for different $p_{\perp,\text{fatjet}}$.

Fig. 6.13 depicts the reconstructed mass of the boosted Z boson. Both in the highly boosted (left) and very highly boosted regimes (right), a peak at the MC truth mass is found over a flat background mass distribution. We consider signal events with $m_{W'} = 3\,\text{TeV}$ (5 TeV) and background events with $\hat{p} \geq 1000\,\text{GeV}$ (2000 GeV) in the $p_\perp \in [1000, 1500]\,\text{GeV}$ ([2000,2500] GeV) bins to approximate realistic $p_\perp$ distributions. No degredation is found in the reconstructed mass when going to high fat jet transverse momenta.

6.3.3 Resonance search with highly boosted gauge bosons

Production of a charged heavy boson, $pp \to W'^{\pm} \to W^{\pm}Z$, as discussed in Sec. 6.3.2 is ideally suited for resonance searches at the LHC. We note that we set the coupling of the $W'WZ$ vertex in accordance with the so-called extended gauge model of Ref. [184] and choose a vector-like coupling to SM quarks. The branching ratio $\text{BR}(W' \to WZ)$ is 3.1–3.2% in the investigated W' mass range $m_{W'} = 3 \cdots 6\,\text{TeV}$. Again we consider two semileptonic scenarios, depending on whether the W boson or the Z boson decays hadronically. In either case, a fat jet is formed with the C/A algorithm with parameters $R = 0.5$ and $p_\perp \geq 800\,\text{GeV}$. The other gauge boson we force to decay leptonically and we assume perfect reconstruction of the visible electrons and muons.

In the first scenario $pp \to W'^{\pm} \to W^{\pm}Z \to (jj)(l^+l^-)$, we apply the `HPTWTagger` to tag and reconstruct the W boson. Identification of the decay of the Z boson into a pair of charged leptons (e^+e^-, $\mu^+\mu^-$) reduces the relevant Standard Model backgrounds to $pp \to Zj$. Here we assume perfect reconstruction of the leptons from the Z boson decay. A fat jet is rejected if close to any of the leptons from the Z decay, $\Delta R(j, \text{lepton}) < 0.6$. We generate background events in bins of generator-level $\hat{p}_\perp$ ranging from 700 GeV to 2.5 TeV, and $\hat{p}_\perp \geq 2.5\,\text{TeV}$.

The heavy W' resonance is reconstructed as the vectorial sum of the tagged W and Z four-momenta. We plot its invariant mass in Fig. 6.14 (l.h.s.), showing expected event rates for an integrated luminosity of $300\,\text{fb}^{-1}$. For all benchmark masses, the signal peaks outnumber the SM background in the respective mass bins.

The second scenario is given by a leptonically decaying W boson, $pp \to W'^{\pm} \to W^{\pm}Z \to (l^{\pm}\nu)(jj)$, and we apply the `HPTZTagger` to the hadronic decay of the Z boson. The relevant Standard Model background is $pp \to W^{\pm}j$ in the highly-boosted regime, again generated in generator-level $\hat{p}_\perp$ bins in the range $[700\,\text{GeV}, 2.5\,\text{TeV}]$ and $\hat{p}_\perp \geq 2.5\,\text{TeV}$.

To recover the four-momentum of the W boson, we assume perfect reconstruction of the charged lepton and obtain the four-momentum of the invisible neutrino as follows. Standard jets (C/A, $R = 0.4$, $p_\perp \geq 30\,\text{GeV}$) are used to determine the missing transverse momentum. Jets in the vicinity of the charged lepton are discarded if $\Delta R_{j,l} < 0.5$. The neutrino transverse momentum is then given by the negative of the charged lepton and summed jet transverse momentum. Imposing zero invariant mass of the neutrino and fixing the W mass $(p_l + p_\nu)^2 = m_W^2$, we can determine the remaining components of the neutrino momentum. Of the two solutions of the quadratic equation for p_z, we choose the one with smaller $\Delta R_{l\nu}$. Due to imperfect knowledge of transverse momentum, there may be no real solution. In this case we simply take the real part as p_z.

Candidate fat jets for Z boson tagging are rejected if not well-separated from the decay products of

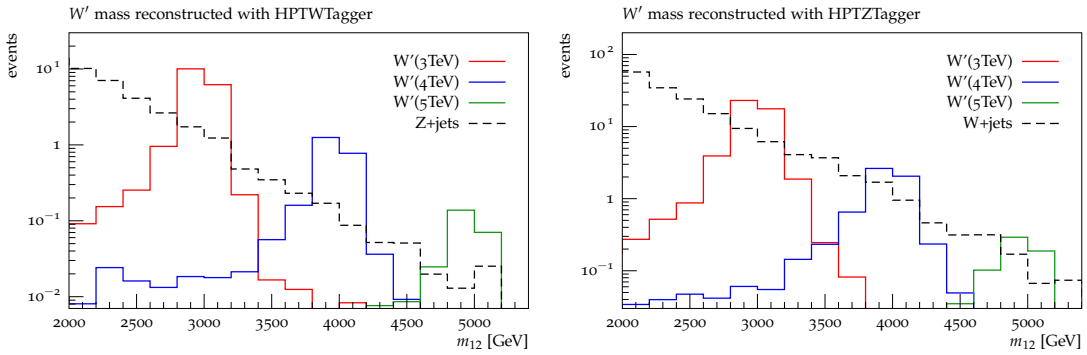


Figure 6.14: Invariant mass of a heavy boson $W'^{\pm} \to W^{\pm}Z$ for three different resonance masses $m_{W'} = 3\,\text{TeV}$ (red), $m_{W'} = 4\,\text{TeV}$ (blue), and $m_{W'} = 5\,\text{TeV}$ (green). In the left panel the hadronically decaying W boson is reconstructed with the `HPTWTagger`, while perfect reconstruction is assumed for the leptonically decaying Z boson. In the right panel the Z boson decays hadronically and is reconstructed with the `HPTZTagger`, while the leptonically decaying W boson is reconstructed as described in the text. The relevant Standard Model backgrounds consist of Z + jets (left) and W + jets (right) respectively, and are depicted as black dashed lines. Event rates are given for pp collisions at $\sqrt{s} = 14\,\text{TeV}$ and an integrated luminosity of $300\,\text{fb}^{-1}$.

Resonance	m_{12} window (TeV)	S/B	$S/\sqrt{B}$	σ(fb)	σ for $S/\sqrt{B} = 2$
$m_{W'} = 3$ TeV	2.9-3.1	10.1	11.9	0.139	0.0233
$m_{W'} = 4$ TeV	3.9-4.1	15.8	5.1	0.0192	0.0075
$m_{W'} = 5$ TeV	4.9-5.1	17.3	1.6	0.00322	0.0040

Table 6.3: Expected event rates for a heavy W' reconstructed with the `HPTWTagger` at $300\,\text{fb}^{-1}$. Note that σ includes the branching fractions BR($W' \to WZ$), BR($W \to$ jets), and BR($Z \to l^+l^-$).

the W boson, $\Delta R(j, \text{lepton/neutrino}) < 0.6$. We then obtain the tagged mass of the heavy W' resonance from the reconstructed W and Z bosons, $m^2_{W'} = (p_W + p_Z)^2$, see Fig. 6.14 (r.h.s.).

As was done in Sec. 6.2.4, we evaluate the required integrated luminosity to exclude the three benchmark resonances using a cut and count analysis. We select the two bins with largest significance ($S/\sqrt{B}$) for the reconstructed W' with the `HPTWTagger` in Tab. 6.3. Using the `HPTZTagger` on the second scenario, we find less significant results, see Tab. 6.4.

Based on the results of the $pp \to W'^{\pm} \to W^{\pm}Z \to (jj)(l^+l^-)$ process, we give the required integrated luminosity to exclude a heavy W' resonance at the LHC depending on the production cross-section in Fig. 6.15.

Resonance	m_{12} window (TeV)	S/B	$S/\sqrt{B}$	σ(fb)	σ for $S/\sqrt{B}=2$
$m_{W'}=3$ TeV	2.9-3.1	3.80	10.6	0.455	0.086
$m_{W'}=4$ TeV	3.9-4.1	2.78	2.92	0.0631	0.043
$m_{W'}=5$ TeV	4.9-5.1	0.087	0.029	0.0108	0.743

Table 6.4: Expected event rates for a heavy W' reconstructed with the `HPTZTagger` at 300 fb^{-1}. Note that σ includes the branching fractions BR($W' \to WZ$), BR($W \to l\nu$), and BR($Z \to$ jets).

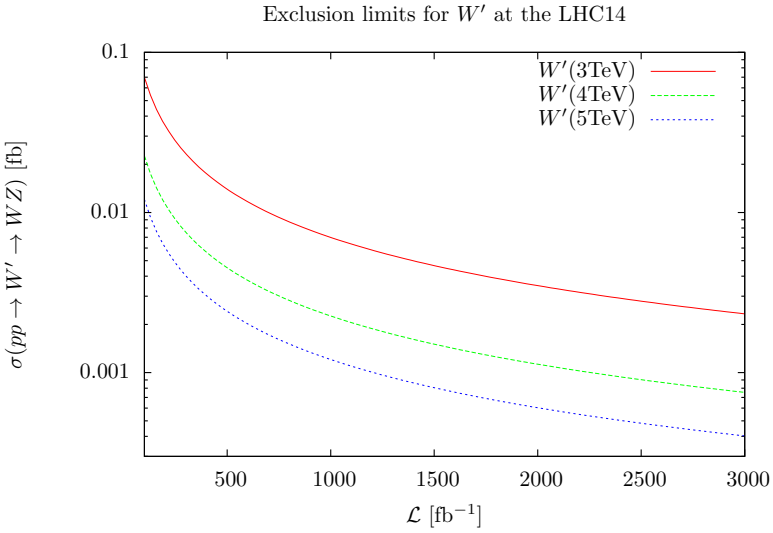


Figure 6.15: Exclusion limits at 90% CL for the process $pp \to W'^{\pm} \to W^{\pm}Z$ at $\sqrt{s} = 14$ TeV. The hadronically decaying W boson is tagged and reconstructed with `HPTWTagger`. Note that σ includes the branching fractions BR($W' \to WZ$), BR($W \to$ jets), and BR($Z \to l^+l^-$).

6.4 Summary and outlook

The reconstruction of heavy resonances is of great importance for the success of the upcoming LHC runs and possible future colliders. When heavy resonances decay into electroweak-scale objects, these objects are highly boosted. The subsequent decay products are then very much collimated and inevitably merge to a single jet, unable to be separately resolved by the detector's calorimeter. For the ATLAS and CMS detectors at the LHC, this angular resolution scale corresponds to heavy resonances with mass $\gtrsim 2\,\text{TeV}$ when the decay products are around the electroweak scale.

We developed dedicated tagging algorithms for the hadronic three-prong decay of the top quark (`HPTTopTagger`) as well as for the hadronic two-prong decay of W and Z bosons (dubbed `HPTWTagger` and `HPTZTagger`). The source code has been made publicly available [F]. The implementation and basic usage are also briefly described in Appendix A.2. Our algorithms show stable efficiencies and kinematic reconstruction up until boosts of several TeV. The taggers combine the good energy resolution of the calorimeter with the very fine spatial resolution of the tracker. As only charged particles leave a signal in the tracking detector, this partial information alone is insufficient. We apply optimized substructure techniques on highly boosted fat jets and find good discrimination power against QCD-initiated jets, which form the dominant background.

Track-based tagging algorithms significantly increase the discovery reach of heavy resonances into the multi-TeV regime. We showed that already in the early stage of run II at the LHC, a heavy Z' gauge boson decaying into a pair of top quarks can be excluded up until $m_{Z'} = 3\,\text{TeV}$, resulting in a much stricter bound than current exclusion limits. A future proton-proton collider with centre-of-mass energy $\sqrt{s} = 100\,\text{TeV}$ is capable of probing heavy resonances of multiples of $\mathcal{O}(10)$ TeV. Applying the `HPTWTagger` and `HPTZTagger` to a search for a heavy charged boson W', we find an exclusion reach up to $m_{W'} = 4\,\text{TeV}$ in the early stage of LHC run II. Probing such large resonance masses is not possible with current tagging algorithms based solely on jets built with calorimeter information.

In fact the general arguments on spatial resolution do not apply exclusively to very heavy resonances. The essential quantity that determines whether or not charged tracks have to be taken into the picture is the mass ratio of the heavy resonance X and the intermediate resonance Y that ultimately decays into the observed jets. Whenever $m_X/m_Y \sim 20$, reconstruction methods relying on calorimeter-based jets break down and track-based observables start to become indispensable. See Appendix 6.A for a demonstration using a rare decay of the Higgs boson.

6.A Tagging highly boosted scalars at the LHC

After the recent discovery of a Higgs boson, i.e. the first electroweak-scale scalar resonance, the scalar sector has become the centre of the focus of both multi-purpose experiments and the theory community. Whether the Higgs boson is part of a minimal or non-minimal Higgs sector remains to be determined. While many models predict several scalar resonances, e.g. 2HDM [185] or general NHDM, their masses are a priori largely undetermined.

To give an example for the benefit of using a track-based tagger for scalar resonances at the LHC with $\sqrt{s} = 14$ TeV, we will focus on the rare prompt decay of the Higgs boson into a Z boson and a

	2 Z bosons	fat jet	$\chi_{jZ} < 15$ GeV	$n_{\text{tracks}} > 1$ and $m_{j_c} < 3$ GeV
$(H \to ZA)Z$	0.513	0.296	0.221	0.131
ZZj	0.497	$1.61 \cdot 10^{-3}$	$4.84 \cdot 10^{-4}$	$5.43 \cdot 10^{-5}$

Table 6.5: Reconstruction efficiencies ϵ for analysis steps described in Sec. 6.A.

CP-odd light scalar $H \to ZA$. The mass of the CP-odd scalar is assumed to be $m_A = 2$ GeV. Hence, A is likely to be highly boosted as $m_H/2 \gg m_A$. We assume A to decay into gluons exclusively. To cope with large backgrounds we study the Higgs boson produced in association with a leptonically decaying Z boson. The process we consider is

$$pp \to HZ \to (ZA)Z \to 4l + gg\,, \tag{6.5}$$

where the gluons are likely to merge into one jet.

We generate signal events using Pythia, and for the background process ZZj we use Sherpa [186], [187]. Including decays of the Z bosons to electrons or muons and imposing $p_{\perp,j} \geq 25$ GeV, we find a leading order cross-section of $\sigma_{ZZj} = 31.69$ fb.

We require exactly four leptons (electrons or muons) with $p_{\perp,l} \geq 10$ GeV. To pair the four leptons to two Z bosons we minimize

$$\chi^2_{ZZ} = \frac{(m_{l_i,l_j} - m_Z)^2}{\Delta_Z^2} + \frac{(m_{l_n,l_m} - m_Z)^2}{\Delta_Z^2} \tag{6.6}$$

where $i \neq j \neq n \neq m$ and $\Delta_Z = 5$ GeV. We require both lepton pairs to be in a mass window of $m_Z \pm 5$ GeV individually. We then remove the leptons from the final state objects and use the remaining objects in $|\eta| < 5$ to cluster C/A jets with $R = 0.4$ and $p_{\perp,j} \geq 30\,\text{GeV}$. Next we need to identify the jet containing the decay products of the CP-odd scalar. Thus, we select this jet by minimizing

$$\chi_{jZ} = \min(|m_{j,Z_1} - m_H|\,, |m_{j,Z_2} - m_H|) \tag{6.7}$$

and veto events where $\chi_{jZ,\text{min}} > 15$ GeV. We find that this requirement selects the jet with smallest distance to the CP-odd scalar efficiently, see Fig. 6.16. For the signal jet we expect a very narrow pencil-like substructure with a high localised energy density, similar to hadronic τ jets. To further remove background contamination we require at least two charged tracks associated[7] with the jet and eventually use all charged tracks as constituents to construct j_c. We veto events if $m_{j_c} > 3$ GeV.

Using the reconstruction efficiencies ϵ in Tab. 6.5, with $\sigma_{HZ} = 883.0$ fb and assuming $\text{BR}(A \to gg) = 1$ we can set a limit $\text{BR}(H \to ZA) < 0.001$ for 100 fb^{-1} integrated luminosity. This Higgs decay mode is usually neglected as it is proportional to $\cos^2(\alpha - \beta)$ in a 2HDM, which is due to Higgs coupling measurements expected to be very small [185]. However, there is no such constraint in general models. Further, the limit derived in this example indicates that it is possible to use the same reconstruction strategy to set an upper limit on exclusive rare Higgs boson decays, e.g. $H \to Z\eta_c$ with $\text{BR}(H \to Z\eta_c) \simeq 1.4 \times 10^{-5}$ [188], $m_{\eta_c} \simeq 2.984$ GeV and $\text{BR}(\eta_c \to gg) \simeq 72\%$.

[7] We use ghost-association, cf. footnote 3 in Sec. 5.3.

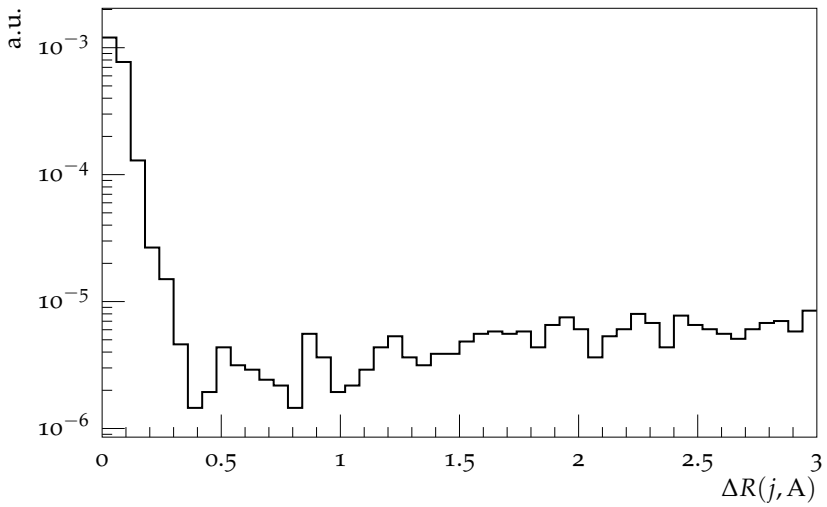


Figure 6.16: Angular distance between the selected jet and the true CP-odd scalar A.

— No disrespect Madam President, but I'm pretty sure it was your idea.

Vivian Wagner

7 Conclusions

Expectations on the Large Hadron Collider are immense, and the first run has indeed been a huge success with the discovery of a Higgs boson. Independent of the outcome of the following runs – whether or not discrepancies to Standard Model predictions are found, whether or not results on new resonances are positive, whether or not unforeseen results appear in the data – the theoretical landscape will change. Many models may be constrained or ruled out completely, while other models may prove to be valid.

At this stage there is strong motivation for new physics around the TeV scale, and also strong motivation that the new states are connected to the top quark and the Higgs sector. Establishing the mechanism of electroweak symmetry breaking and the Higgs sector has been the leading motivation to build the LHC in the first place. Efficient identification of top quarks is a prime tool to gain access to the relevant processes. As tops emerging from new physics processes are likely to be highly energetic, boosted top taggers have come into focus. The commonly used state-of-the-art algorithms construct a large-radius "fat" jet and investigate its substructure to discriminate between signal and background.

Progressing towards the energy frontier to make maximum use of the LHC, there are, however, scenarios where these tools start to find their limitations. When new heavy particles decay into Standard Model resonances, the final state may be populated with radiation from several top quarks, Higgs or gauge bosons. As an example, we investigated pair-production of vectorlike top partners with mass around 1 TeV. Their subsequent decay into a top quark and a Higgs boson each leads to a very busy final state of boosted top quarks and Higgs bosons, which cannot be easily separated into individual fat jets. In this dissertation we suggested a new jet clustering algorithm that harnesses the advantages of substructure analyses without constructing the intermediate fat jet. Its key feature is a terminating veto that will halt further clustering of two prongs whose combined jet mass is substantially larger than the individual

invariant masses. This "mass jump" criterion effectively clusters jets with variable radii and helps to separately resolve nearby jets. Using the mass-jump algorithm, we find that our analysis of this process is superior to common jet algorithms as well as a similar analysis based on sophisticated fat jet-based taggers.

The LHC is also capable to produce very heavy new gauge bosons that can decay into highly boosted Standard Model resonances. Above a certain transverse momentum – typically around 1 TeV for top quarks and 500 MeV for W and Z gauge bosons – however, their decay products are so closely collimated that the spatial resolution of the detector's calorimeters is insufficient to resolve the relevant substructure. In this regime it becomes necessary to take charged tracks into account and combine their superior spatial resolution with the good energy resolution of the calorimeters. The resulting algorithm developed and investigated in this dissertation is dubbed `HPTTopTagger`. We also proposed similar taggers for electroweak gauge bosons, the `HPTWTagger` and `HPTZTagger`. In realistic analyses we found that these algorithms can significantly extend the discovery reach for new physics into the multi-TeV regime at the LHC and also future hadron colliders.

The LHC is pushing forward the energy frontier to a level where the pursuit for the correct theory of Nature will claim many victims. For the first time, heavy TeV-scale particles could be produced in an experiment. To make maximum use of this unique opportunity, searches have to be conducted for their characteristic final states, which may contain a multitude of Standard Model resonances or highly boosted particles. This dissertation aims to provide a starting point for these very final states. Hopefully the tools we developed with those scenarios in mind will prove useful during this quest.

— I'll be Mr. Purple.

Mr. Pink

A Implementation and usage

A.1 Jet clustering with a terminating veto

The code for jet clustering with a terminating veto is written in `C++` and implemented as a plugin for the FastJet contribution package. The implementation is based on the algorithm developed in Chapter 4. It is publicly available from source [E].

The plugin is dubbed `ClusteringVetoPlugin` and contains the identically named **class** ClusteringVetoPlugin. A minimal usage example for the default mass-jump setup is the following.

```
#include <vector>
#include "fastjet/PseudoJet.hh"
#include "fastjet/contrib/ClusteringVetoPlugin.hh"

int main {
  // read in input particles //
  std::vector<fastjet::PseudoJet> event;
  read_event( event ); // defined elsewhere

  // set mass-jump clustering parameters //
  double mu(30.), theta(0.7), max_r(1.0);
```

```
  fastjet::contrib::ClusteringVetoPlugin mj_plugin
    ( mu, theta, max_r, fastjet::contrib::ClusteringVetoPlugin::CALIKE );
  fastjet::JetDefinition jet_def( &mj_plugin );
  fastjet::ClusterSequence clust_seq( event, jet_def );

  // inclusive jets //
  double min_pt(25.);
  std::vector<fastjet::PseudoJet> jets =
    fastjet::sorted_by_pt( clust_seq.inclusive_jets(min_pt) );
}
```

The currently available options for the distance metric, Eq. (3.1), are collected in the enumeration fastjet :: contrib :: ClusteringVetoPlugin::ClusterType. Its identifiers are

- CALIKE
- KTLIKE
- AKTLIKE

To allow for maximum flexibility, the user can specify any custom veto function. The user-defined veto function takes the following form,

```
fastjet::contrib::ClusteringVetoPlugin::VetoResult user_veto
( const fastjet::PseudoJet&, const fastjet::PseudoJet& );
```

The enumeration fastjet :: contrib :: ClusteringVetoPlugin::VetoResult contains the identifiers

- CLUSTER
  Cluster the two prongs. In the mass-jump setup this corresponds to the case where the combined jet mass is below the threshold μ, cf. the first sentence in bullet point 2 of Sec. 4.2.2.
- VETO
  Veto the merging step and label both prongs passive. In the mass-jump setup this corresponds to the case where a mass jump could be identified, cf. second sentence in bullet point 2 of Sec. 4.2.2.
- NOVETO
  Do not veto the merging step, but active-passive vetos may have to be checked. In the mass-jump setup this corresponds to the case where the combined jet mass is above the threshold μ, but the mass-jump condition is not fulfilled, leading to bullet points 3 and 4 of Sec. 4.2.2.

A simple example of a user-defined veto function is modified mass-jump clustering with the jet mass threshold μ replaced by a minimum angular separation, e.g.

```
fastjet::contrib::ClusteringVetoPlugin::VetoResult user_veto_function
( const fastjet::PseudoJet& j1, const fastjet::PseudoJet& j2 ) {
  // DeltaR instead of jet mass threshold
  if( j1.delta_R(j2) < 0.3 )
    return fastjet::contrib::ClusteringVetoPlugin::CLUSTER;
  // mass-jump veto with theta=0.5
  else if( 0.5 * (j1+j2).m() > max( j1.m(), j2.m() ) )
    return fastjet::contrib::ClusteringVetoPlugin::VETO;
  // no veto, but algorithm may need to check active-passive veto
  return fastjet::contrib::ClusteringVetoPlugin::NOVETO;
}
```

The user-defined veto function can be set with

```
mj_plugin.set_veto_function( user_veto_function );
```

Note that the constructor of the ClusteringVetoPlugin class requires the mass-jump parameters μ and θ, which can be set to arbitrary values.

A.2 High-$p_\perp$ tagging algorithms

Tagging algorithms designed for the highly-boosted energy regime, `HPTTopTagger`, `HPTWTagger`, and `HPTZTagger`, were developed in Chapter 6. They are implemented in `C++` and available from source [F].

A.2.1 `HPTTopTagger`

All routines are collected in the **class** HPTTopTagger. The following describes a minimal usage example.

```
#include <vector>
#include "fastjet/PseudoJet.hh"
#include "HPTTopTagger.hh"

int main {
  // fat jet and track-based fat jet //
  fastjet::PseudoJet jet, c_jet;
  fastjet::ClusterSequence clust_seq, c_clust_seq;
  // initialize these objects here

  // run the HPTTopTagger //
  HPTTopTagger hpttoptag( c_clust_seq, c_jet, clust_seq, jet );
  hpttoptag.run_tagger();
```

```
  if( hpttoptag.is_maybe_top() ) {
    ; // top candidate within mass window is found

    if( hpttoptag.is_tagged() ) {
      ; // top candidate satisfies all cuts
    }
  }
}
```

The constructor takes the following arguments,

1. the fastjet::ClusterSequence of the track/e-cal-based fat jet
2. the corresponding fastjet::PseudoJet
3. the fastjet::ClusterSequence of the calorimeter-based fat jet
4. the corresponding fastjet::PseudoJet

and two optional arguments,

5. top quark mass
6. W boson mass.

Among other objects, the top candidate can be obtained as a fastjet::PseudoJet by calling top_candidate(). Information about the candidate is written to std::cout with the get_info() routine. The setup including all relevant parameters is written via the get_setting() routine. We refer to the documented source code for all available options.

In Chapter 6 we suggest to use ghost association (recall footnote 3 in Sec. 5.3 for a definition) to identify the charged tracks that correspond to a given (calorimeter-based) fat jet. These tracks form the track-based fat jet. The following depicts one possible implementation.

```
// necessary objects to be initialized elsewhere //
// calorimeter-based fat jet
fastjet::JetDefinition jet_def;
fastjet::ClusterSequence clust_seq;
fastjet::PseudoJet jet;
// charged tracks
vector<fastjet::PseudoJet> tracks;

// create ghost tracks //
vector<fastjet::PseudoJet> ghost_tracks;
foreach( fastjet::PseudoJet const p, tracks ) {
  fastjet::PseudoJet pp;
```

```
  pp.reset_PtYPhiM( 0.00001, p.eta(), p.phi(), 0.00000001 );
  pp.set_user_index( ghost_tracks.size() );
  ghost_tracks.push_back( pp );
}

// associate ghosts to fat jet //
vector<fastjet::PseudoJet> jet_constituents =
  clust_seq.constituents( jet );
jet_constituents.insert( jet_constituents.end(),
                         ghost_tracks.begin(), ghost_tracks.end() );

fastjet::ClusterSequence clust_temp( jet_constituents, jet_def );
fastjet::PseudoJet jet_withtracks =
  fastjet::sorted_by_pt( clust_temp.inclusive_jets() )[0];

// extract constituents
vector<fastjet::PseudoJet> ghost_constituents =
  clust_temp.constituents( jet_withtracks );

// extract ghosts and restore tracks
vector<fastjet::PseudoJet> charged_constituents;
for( unsigned j=0; j<ghost_constituents.size(); ++j ) {
  if( fabs(ghost_constituents[j].pt() - 0.00001) < 1e-8 ) {
    charged_constituents.push_back
      ( tracks[ ghost_constituents[j].user_index() ] );
  }
 }
if( charged_constituents.size() == 0 )
  continue;

// create track-based fat jet //
fastjet::JetDefinition c_jet_def( fastjet::cambridge_algorithm,
                                  2.*fatjet_R_ );
fastjet::ClusterSequence c_clust_seq( charged_constituents, c_jet_def );
fastjet::PseudoJet c_jet =
  fastjet::sorted_by_pt( c_clust_seq.inclusive_jets() )[0];
```

A.2.2 `HPTWTagger` and `HPTZTagger`

The routines of the tagging algorithms for electroweak gauge bosons, `HPTWTagger` and `HPTZTagger`, closely follow those of the `HPTTopTagger`. Their classes are dubbed HPTWTagger and HPTZTagger,

respectively. Both derive from a common class labelled HPTEWBTagger.

The naming of all routines is according to the corresponding ones of the `HPTTopTagger` implementation, in particular the calls

- run_tagger()
- is_maybe_w()
- is_tagged()
- w_candidate()

are implemented, as well as their corresponding calls in the HPTZTagger class.

By default, all parameters are set in accordance to working point 1 (cf. Sec. 6.3.1). Switching to working point 2 is possible with the routine set_working_point_2().

Bibliography

[1] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl.Phys.*, vol. 22, pp. 579–588, 1961. DOI: 10.1016/0029-5582(61)90469-2.

[2] S. Weinberg, "A Model of Leptons," *Phys.Rev.Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: 10.1103/PhysRevLett.19.1264.

[3] A. Salam, "Weak and Electromagnetic Interactions," *Conf.Proc.*, vol. C680519, pp. 367–377, 1968.

[4] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys.Lett.*, vol. 8, pp. 214–215, 1964. DOI: 10.1016/S0031-9163(64)92001-3.

[5] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys.Lett.*, vol. B47, pp. 365–368, 1973. DOI: 10.1016/0370-2693(73)90625-4.

[6] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories," *Phys.Rev.Lett.*, vol. 30, pp. 1343–1346, 1973. DOI: 10.1103/PhysRevLett.30.1343.

[7] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?" *Phys.Rev.Lett.*, vol. 30, pp. 1346–1349, 1973. DOI: 10.1103/PhysRevLett.30.1346.

[8] G. 't Hooft, "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking," *NATO Sci.Ser.B*, vol. 59, p. 135, 1980.

[9] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys.Lett.*, vol. 12, pp. 132–133, 1964. DOI: 10.1016/0031-9163(64)91136-9.

[10] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys.Rev.Lett.*, vol. 13, pp. 508–509, 1964. DOI: 10.1103/PhysRevLett.13.508.

[11] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys.Rev.Lett.*, vol. 13, pp. 321–323, 1964. DOI: 10.1103/PhysRevLett.13.321.

[12] G. Guralnik, C. Hagen, and T. Kibble, "Global Conservation Laws and Massless Particles," *Phys.Rev.Lett.*, vol. 13, pp. 585–587, 1964. DOI: 10.1103/PhysRevLett.13.585.

[13] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[14] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys.Lett.*, vol. B716, pp. 30–61, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[15] S. P. Martin, "A Supersymmetry primer," *Adv.Ser.Direct.High Energy Phys.*, vol. 21, pp. 1–153, 2010. DOI: 10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph].

[16] M. Schmaltz and D. Tucker-Smith, "Little Higgs review," *Ann.Rev.Nucl.Part.Sci.*, vol. 55, pp. 229–270, 2005. DOI: 10.1146/annurev.nucl.55.090704.151502. arXiv: hep-ph/0502182 [hep-ph].

[17] D. E. Morrissey, T. Plehn, and T. M. Tait, "Physics searches at the LHC," *Phys.Rept.*, vol. 515, pp. 1–113, 2012. DOI: 10.1016/j.physrep.2012.02.007. arXiv: 0912.3259 [hep-ph].

[18] A. Abdesselam, E. B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, *et al.*, "Boosted objects: A Probe of beyond the Standard Model physics," *Eur.Phys.J.*, vol. C71, p. 1661, 2011. DOI: 10.1140/epjc/s10052-011-1661-y. arXiv: 1012.5412 [hep-ph].

[19] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, *et al.*, "Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks," *J.Phys.*, vol. G39, p. 063001, 2012. DOI: 10.1088/0954-3899/39/6/063001. arXiv: 1201.0008 [hep-ph].

[20] A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann, *et al.*, "Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012," *Eur.Phys.J.*, vol. C74, no. 3, p. 2792, 2014. DOI: 10.1140/epjc/s10052-014-2792-8. arXiv: 1311.2708 [hep-ex].

[21] D. Adams, A. Arce, L. Asquith, M. Backovic, T. Barillari, *et al.*, "Towards an Understanding of the Correlations in Jet Substructure," 2015. arXiv: 1504.00679 [hep-ph].

[22] M. H. Seymour, "Searches for new particles using cone and cluster jet algorithms: A Comparative study," *Z.Phys.*, vol. C62, pp. 127–138, 1994. DOI: 10.1007/BF01559532.

[23] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC," *Phys.Rev.Lett.*, vol. 100, p. 242001, 2008. DOI: 10.1103/PhysRevLett.100.242001. arXiv: 0802.2470 [hep-ph].

[24] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, "Stop Reconstruction with Tagged Tops," *JHEP*, vol. 1010, p. 078, 2010. DOI: 10.1007/JHEP10(2010)078. arXiv: 1006.2833 [hep-ph].

[25] S. Schaetzel and M. Spannowsky, "Tagging highly boosted top quarks," *Phys.Rev.*, vol. D89, no. 1, p. 014007, 2014. DOI: 10.1103/PhysRevD.89.014007. arXiv: 1308.0540 [hep-ph].

[26] T. Cheng and L. Li, "GAUGE THEORY OF ELEMENTARY PARTICLE PHYSICS," 1984.

[27] F. Mandl and G. Shaw, "QUANTUM FIELD THEORY," 1985.

[28] L. Ryder, "QUANTUM FIELD THEORY," 1985.

[29] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory," 1995.

[30] G. L. Kane, "Physics beyond the standard model," *Sci.Am.*, vol. 288N6, pp. 68–75, 2003. DOI: 10.1038/scientificamerican0603-68.

[31] H. Georgi, S. Glashow, M. Machacek, and D. V. Nanopoulos, "Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions," *Phys.Rev.Lett.*, vol. 40, p. 692, 1978. DOI: 10.1103/PhysRevLett.40.692.

[32] J. Baglio and A. Djouadi, "Predictions for Higgs production at the Tevatron and the associated uncertainties," *JHEP*, vol. 1010, p. 064, 2010. DOI: 10.1007/JHEP10(2010)064. arXiv: 1003.4266 [hep-ph].

[33] J. Baglio and A. Djouadi, "Higgs production at the lHC," *JHEP*, vol. 1103, p. 055, 2011. DOI: 10.1007/JHEP03(2011)055. arXiv: 1012.0530 [hep-ph].

[34] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, *et al.*, "Scalar and pseudoscalar Higgs production in association with a top-antitop pair," *Phys.Lett.*, vol. B701, pp. 427–433, 2011. DOI: 10.1016/j.physletb.2011.06.012. arXiv: 1104.5613 [hep-ph].

[35] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, "Standard Model Higgs boson production in association with a top anti-top pair at NLO with parton showering," *Europhys.Lett.*, vol. 96, p. 11 001, 2011. DOI: 10.1209/0295-5075/96/11001. arXiv: 1108.0387 [hep-ph].

[36] C. Degrande, J. Gerard, C. Grojean, F. Maltoni, and G. Servant, "Probing Top-Higgs Non-Standard Interactions at the LHC," *JHEP*, vol. 1207, p. 036, 2012. DOI: 10.1007/JHEP07(2012)036,10.1007/JHEP03(2013)032. arXiv: 1205.1065 [hep-ph].

[37] D. Curtin, J. Galloway, and J. G. Wacker, "Measuring the $t\bar{t}h$ coupling from same-sign dilepton $+2b$ measurements," *Phys.Rev.*, vol. D88, no. 9, p. 093 006, 2013. DOI: 10.1103/PhysRevD.88.093006. arXiv: 1306.5695 [hep-ph].

[38] P. Agrawal, S. Bandyopadhyay, and S. P. Das, "Multilepton Signatures of the Higgs Boson through its Production in Association with a Top-quark Pair," *Phys.Rev.*, vol. D88, no. 9, p. 093 008, 2013. DOI: 10.1103/PhysRevD.88.093008. arXiv: 1308.3043 [hep-ph].

[39] J. Adelman, A. Loginov, P. Tipton, and J. Vasquez, "Study of $t\bar{t}H$ ($H \to \mu^+\bar{\mu}^-$) in the three lepton channel at $\sqrt{s} = 14$ TeV; A Snowmass white paper," 2013. arXiv: 1310.1132 [hep-ex].

[40] S. Biswas, R. Frederix, E. Gabrielli, and B. Mele, "Enhancing the $t\bar{t}H$ signal through top-quark spin polarization effects at the LHC," *JHEP*, vol. 1407, p. 020, 2014. DOI: 10.1007/JHEP07(2014)020. arXiv: 1403.1790 [hep-ph].

[41] T. A. collaboration, "Search for ttH production in the H -¿ yy channel at sqrt(s) = 8 TeV with the ATLAS detector," 2013.

[42] T. A. collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC," 2014.

[43] G. Aad *et al.*, "Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector," *Phys.Lett.*, vol. B740, pp. 222–242, 2015. DOI: 10.1016/j.physletb.2014.11.049. arXiv: 1409.3122 [hep-ex].

[44] T. A. collaboration, "Search for the associated production of the Higgs boson with a top quark pair in multi-lepton final states with the ATLAS detector," 2015.

[45] G. Aad *et al.*, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," 2015. arXiv: 1503.05066 [hep-ex].

[46] S. Chatrchyan *et al.*, "Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC," *JHEP*, vol. 1305, p. 145, 2013. DOI: 10.1007/JHEP05(2013)145. arXiv: 1303.0763 [hep-ex].

[47] V. Khachatryan *et al.*, "Search for the associated production of the Higgs boson with a top-quark pair," *JHEP*, vol. 1409, p. 087, 2014. DOI: 10.1007/JHEP09(2014)087,10.1007/JHEP10(2014)106. arXiv: 1408.1682 [hep-ex].

[48] V. Khachatryan *et al.*, "Search for a standard model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method," *Eur.Phys.J.*, vol. C75, no. 6, p. 251, 2015. DOI: 10.1140/epjc/s10052-015-3454-1. arXiv: 1502.02485 [hep-ex].

[49] F. Maltoni, "Extending the lhc reach: the search for new interactions," KAERU conference, 2015, [Online]. Available: http://member.ipmu.jp/kaeru/.

[50] I. I. Bigi, Y. L. Dokshitzer, V. A. Khoze, J. H. Kuhn, and P. M. Zerwas, "Production and Decay Properties of Ultraheavy Quarks," *Phys.Lett.*, vol. B181, p. 157, 1986. DOI: 10.1016/0370-2693(86)91275-X.

[51] H. Georgi and A. Pais, "Calculability and Naturalness in Gauge Theories," *Phys.Rev.*, vol. D10, p. 539, 1974. DOI: 10.1103/PhysRevD.10.539.

[52] H. Georgi and A. Pais, "Vacuum Symmetry and the PseudoGoldstone Phenomenon," *Phys.Rev.*, vol. D12, p. 508, 1975. DOI: 10.1103/PhysRevD.12.508.

[53] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, "Electroweak symmetry breaking from dimensional deconstruction," *Phys.Lett.*, vol. B513, pp. 232–240, 2001. DOI: 10.1016/S0370-2693(01)00741-9. arXiv: hep-ph/0105239 [hep-ph].

[54] N. Arkani-Hamed, A. G. Cohen, T. Gregoire, and J. G. Wacker, "Phenomenology of electroweak symmetry breaking from theory space," *JHEP*, vol. 0208, p. 020, 2002. arXiv: hep-ph/0202089 [hep-ph].

[55] M. Schmaltz, "Physics beyond the standard model (theory): Introducing the little Higgs," *Nucl.Phys.Proc.Suppl.*, vol. 117, pp. 40–49, 2003. DOI: 10.1016/S0920-5632(03)01409-9. arXiv: hep-ph/0210415 [hep-ph].

[56] D. E. Kaplan and M. Schmaltz, "The Little Higgs from a simple group," *JHEP*, vol. 0310, p. 039, 2003. DOI: 10.1088/1126-6708/2003/10/039. arXiv: hep-ph/0302049 [hep-ph].

[57] M. Schmaltz, "The Simplest little Higgs," *JHEP*, vol. 0408, p. 056, 2004. DOI: 10.1088/1126-6708/2004/08/056. arXiv: hep-ph/0407143 [hep-ph].

[58] N. Arkani-Hamed, A. Cohen, E. Katz, A. Nelson, T. Gregoire, *et al.*, "The Minimal moose for a little Higgs," *JHEP*, vol. 0208, p. 021, 2002. DOI: 10.1088/1126-6708/2002/08/021. arXiv: hep-ph/0206020 [hep-ph].

[59] M. J. Dugan, H. Georgi, and D. B. Kaplan, "Anatomy of a Composite Higgs Model," *Nucl.Phys.*, vol. B254, p. 299, 1985. DOI: 10.1016/0550-3213(85)90221-4.

[60] N. Arkani-Hamed, A. Cohen, E. Katz, and A. Nelson, "The Littlest Higgs," *JHEP*, vol. 0207, p. 034, 2002. DOI: 10.1088/1126-6708/2002/07/034. arXiv: hep-ph/0206021 [hep-ph].

[61] S. Weinberg, "Implications of Dynamical Symmetry Breaking," *Phys.Rev.*, vol. D13, pp. 974–996, 1976. DOI: 10.1103/PhysRevD.13.974.

[62] L. Susskind, "Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory," *Phys.Rev.*, vol. D20, pp. 2619–2625, 1979. DOI: 10.1103/PhysRevD.20.2619.

[63] K. Lane, "Two lectures on technicolor," 2002. arXiv: hep-ph/0202255 [hep-ph].

[64] C. T. Hill and E. H. Simmons, "Strong dynamics and electroweak symmetry breaking," *Phys.Rept.*, vol. 381, pp. 235–402, 2003. DOI: 10.1016/S0370-1573(03)00140-6. arXiv: hep-ph/0203079 [hep-ph].

[65] R. Shrock, "Some recent results on models of dynamical electroweak symmetry breaking," pp. 227–241, 2007. arXiv: hep-ph/0703050 [HEP-PH].

[66] R. M. Harris, C. T. Hill, and S. J. Parke, "Cross-section for topcolor Z-prime(t) decaying to t anti-t: Version 2.6," 1999. arXiv: hep-ph/9911288 [hep-ph].

[67] R. M. Harris and S. Jain, "Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop," *Eur.Phys.J.*, vol. C72, p. 2072, 2012. DOI: 10.1140/epjc/s10052-012-2072-4. arXiv: 1112.4928 [hep-ph].

[68] S. R. Coleman and J. Mandula, "All Possible Symmetries of the S Matrix," *Phys.Rev.*, vol. 159, pp. 1251–1256, 1967. DOI: 10.1103/PhysRev.159.1251.

[69] R. Haag, J. T. Lopuszanski, and M. Sohnius, "All Possible Generators of Supersymmetries of the s Matrix," *Nucl.Phys.*, vol. B88, p. 257, 1975. DOI: 10.1016/0550-3213(75)90279-5.

[70] P. Bechtle, K. Desch, H. K. Dreiner, M. Hamer, M. Kramer, *et al.*, "How alive is constrained SUSY really?," 2014. arXiv: 1410.6035 [hep-ph].

[71] T. Moroi and Y. Okada, "Upper bound of the lightest neutral Higgs mass in extended supersymmetric Standard Models," *Phys.Lett.*, vol. B295, pp. 73–78, 1992. DOI: 10.1016/0370-2693(92)90091-H.

[72] K. Babu, I. Gogoladze, and C. Kolda, "Perturbative unification and Higgs boson mass bounds," 2004. arXiv: hep-ph/0410085 [hep-ph].

[73] S. P. Martin, "Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry," *Phys.Rev.*, vol. D81, p. 035 004, 2010. DOI: 10.1103/PhysRevD.81.035004. arXiv: 0910.2732 [hep-ph].

[74] M. Endo, K. Hamaguchi, S. Iwamoto, and N. Yokozaki, "Higgs Mass and Muon Anomalous Magnetic Moment in Supersymmetric Models with Vector-Like Matters," *Phys.Rev.*, vol. D84, p. 075 017, 2011. DOI: 10.1103/PhysRevD.84.075017. arXiv: 1108.3071 [hep-ph].

[75] T. Moroi, R. Sato, and T. T. Yanagida, "Extra Matters Decree the Relatively Heavy Higgs of Mass about 125 GeV in the Supersymmetric Model," *Phys.Lett.*, vol. B709, pp. 218–221, 2012. DOI: 10.1016/j.physletb.2012.02.012. arXiv: 1112.3142 [hep-ph].

[76] M. Endo, K. Hamaguchi, K. Ishikawa, S. Iwamoto, and N. Yokozaki, "Gauge Mediation Models with Vectorlike Matters at the LHC," *JHEP*, vol. 1301, p. 181, 2013. DOI: 10.1007/JHEP01(2013)181. arXiv: 1212.3935 [hep-ph].

[77] T. A. collaboration, *Exotics public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults, retrieved on June 11, 2015.

[78] C. Collaboration, *Beyond-two-generations public results*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G, retrieved on June 11, 2015.

[79] F. Abe *et al.*, "Observation of top quark production in $\bar{p}p$ collisions," *Phys.Rev.Lett.*, vol. 74, pp. 2626–2631, 1995. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex].

[80] S. Abachi *et al.*, "Observation of the top quark," *Phys.Rev.Lett.*, vol. 74, pp. 2632–2637, 1995. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003 [hep-ex].

[81] M. Vos, "High pT hadronic top quark identification. Part II: The lifetime signature," 2008.

[82] J. Thaler and L.-T. Wang, "Strategies to Identify Boosted Tops," *JHEP*, vol. 0807, p. 092, 2008. DOI: 10.1088/1126-6708/2008/07/092. arXiv: 0806.0023 [hep-ph].

[83] K. Rehermann and B. Tweedie, "Efficient Identification of Boosted Semileptonic Top Quarks at the LHC," *JHEP*, vol. 1103, p. 059, 2011. DOI: 10.1007/JHEP03(2011)059. arXiv: 1007.2221 [hep-ph].

[84] T. Plehn, M. Spannowsky, and M. Takeuchi, "Boosted Semileptonic Tops in Stop Decays," *JHEP*, vol. 1105, p. 135, 2011. DOI: 10.1007/JHEP05(2011)135. arXiv: 1102.0557 [hep-ph].

[85] G. Kasieczka, T. Plehn, T. Schell, T. Strebler, and G. P. Salam, "Resonance Searches with an Updated Top Tagger," 2015. arXiv: 1503.05921 [hep-ph].

[86] G. Brooijmans, "High pT hadronic top quark identification. Part I: Jet mass and YSplitter," 2008.

[87] "Reconstruction of High Mass $t\bar{t}$ Resonances in the Lepton+Jets Channel," 2009.

[88] "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging," 2009.

[89] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks," *Phys.Rev.Lett.*, vol. 101, p. 142 001, 2008. DOI: 10.1103/PhysRevLett.101.142001. arXiv: 0806.0848 [hep-ph].

[90] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness," *JHEP*, vol. 1103, p. 015, 2011. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[91] J.-H. Kim, "Rest Frame Subjet Algorithm With SISCone Jet For Fully Hadronic Decaying Higgs Search," *Phys.Rev.*, vol. D83, p. 011 502, 2011. DOI: 10.1103/PhysRevD.83.011502. arXiv: 1011.1493 [hep-ph].

[92] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Techniques for improved heavy particle searches with jet substructure," *Phys.Rev.*, vol. D80, p. 051 501, 2009. DOI: 10.1103/PhysRevD.80.051501. arXiv: 0903.5081 [hep-ph].

[93] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches," *Phys.Rev.*, vol. D81, p. 094 023, 2010. DOI: 10.1103/PhysRevD.81.094023. arXiv: 0912.0033 [hep-ph].

[94] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming," *JHEP*, vol. 1002, p. 084, 2010. DOI: 10.1007/JHEP02(2010)084. arXiv: 0912.1342 [hep-ph].

[95] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, *et al.*, "HERWIG 6.5 release note," 2002. arXiv: hep-ph/0210213 [hep-ph].

[96] D. Lopez Mateos, "Measurement of multi-jet production cross section at a center-of-mass energy of 7 TeV at the large hadron collider with the ATLAS detector,"

[97] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," 2009.

[98] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. Webber, "New clustering algorithm for multi - jet cross-sections in e+ e- annihilation," *Phys.Lett.*, vol. B269, pp. 432–438, 1991. DOI: 10.1016/0370-2693(91)90196-W.

[99] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions," *Nucl.Phys.*, vol. B406, pp. 187–224, 1993. DOI: 10.1016/0550-3213(93)90166-M.

[100] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys.Rev.*, vol. D48, pp. 3160–3166, 1993. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[101] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms," *JHEP*, vol. 9708, p. 001, 1997. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[102] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering," 1998. arXiv: hep-ph/9907280 [hep-ph].

[103] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP*, vol. 0804, p. 063, 2008. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[104] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. D. Elvira, K. Frame, *et al.*, "Run II jet physics," pp. 47–77, 2000. arXiv: hep-ex/0005012 [hep-ex].

[105] G. F. Sterman and S. Weinberg, "Jets from Quantum Chromodynamics," *Phys.Rev.Lett.*, vol. 39, p. 1436, 1977. DOI: 10.1103/PhysRevLett.39.1436.

[106] M. G. Albrow *et al.*, "Tevatron-for-LHC Report of the QCD Working Group," 2006. arXiv: hep-ph/0610012 [hep-ph].

[107] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm," *JHEP*, vol. 0705, p. 086, 2007. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph].

[108] D. Bertolini, T. Chan, and J. Thaler, "Jet Observables Without Jet Algorithms," *JHEP*, vol. 1404, p. 013, 2014. DOI: 10.1007/JHEP04(2014)013. arXiv: 1310.7584 [hep-ph].

[109] H. Georgi, "A Simple Alternative to Jet-Clustering Algorithms," 2014. arXiv: 1408.1161 [hep-ph].

[110] S.-F. Ge, "The Georgi Algorithms of Jet Clustering," *JHEP*, vol. 1505, p. 066, 2015. DOI: 10.1007/JHEP05(2015)066. arXiv: 1408.3823 [hep-ph].

[111] M. Dasgupta, L. Magnea, and G. P. Salam, "Non-perturbative QCD effects in jets at hadron colliders," *JHEP*, vol. 0802, p. 055, 2008. DOI: 10.1088/1126-6708/2008/02/055. arXiv: 0712.3014 [hep-ph].

[112] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez, "Quantifying the performance of jet definitions for kinematic reconstruction at the LHC," *JHEP*, vol. 0812, p. 032, 2008. DOI: 10.1088/1126-6708/2008/12/032. arXiv: 0810.1304 [hep-ph].

[113] G. Soyez, "Optimal jet radius in kinematic dijet reconstruction," *JHEP*, vol. 1007, p. 075, 2010. DOI: 10.1007/JHEP07(2010)075. arXiv: 1006.3634 [hep-ph].

[114] M. Cacciari, G. P. Salam, and S. Sapeta, "On the characterisation of the underlying event," *JHEP*, vol. 1004, p. 065, 2010. DOI: 10.1007/JHEP04(2010)065. arXiv: 0912.4926 [hep-ph].

[115] Z. Marshall, "Simulation of Pile-up in the ATLAS Experiment," *J.Phys.Conf.Ser.*, vol. 513, p. 022 024, 2014. DOI: 10.1088/1742-6596/513/2/022024.

[116] C. Collaboration, "Pileup Removal Algorithms," 2014.

[117] H.-n. Li, Z. Li, and C.-P. Yuan, "QCD resummation for light-particle jets," *Phys.Rev.*, vol. D87, p. 074 025, 2013. DOI: 10.1103/PhysRevD.87.074025. arXiv: 1206.1344 [hep-ph].

[118] M. Dasgupta, K. Khelifa-Kerfa, S. Marzani, and M. Spannowsky, "On jet mass distributions in Z+jet and dijet processes at the LHC," *JHEP*, vol. 1210, p. 126, 2012. DOI: 10.1007/JHEP10(2012)126. arXiv: 1207.1640 [hep-ph].

[119] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure," *JHEP*, vol. 1309, p. 029, 2013. DOI: 10.1007/JHEP09(2013)029. arXiv: 1307.0007 [hep-ph].

[120] M. Dasgupta, A. Fregoso, S. Marzani, and A. Powling, "Jet substructure with analytical methods," *Eur.Phys.J.*, vol. C73, no. 11, p. 2623, 2013. DOI: 10.1140/epjc/s10052-013-2623-3. arXiv: 1307.0013 [hep-ph].

[121] M. Dasgupta and G. P. Salam, "Event shapes in e+ e- annihilation and deep inelastic scattering," *J.Phys.*, vol. G30, R143, 2004. DOI: 10.1088/0954-3899/30/5/R01. arXiv: hep-ph/0312283 [hep-ph].

[122] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness," *JHEP*, vol. 1202, p. 093, 2012. DOI: 10.1007/JHEP02(2012)093. arXiv: 1108.2701 [hep-ph].

[123] D. E. Soper and M. Spannowsky, "Combining subjet algorithms to enhance ZH detection at the LHC," *JHEP*, vol. 1008, p. 029, 2010. DOI: 10.1007/JHEP08(2010)029. arXiv: 1005.0417 [hep-ph].

[124] D. E. Soper and M. Spannowsky, "Finding top quarks with shower deconstruction," *Phys.Rev.*, vol. D87, p. 054012, 2013. DOI: 10.1103/PhysRevD.87.054012. arXiv: 1211.3140 [hep-ph].

[125] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, "Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure," *Phys.Rev.Lett.*, vol. 108, p. 182003, 2012. DOI: 10.1103/PhysRevLett.108.182003. arXiv: 1201.1914 [hep-ph].

[126] S. D. Ellis, A. Hornig, D. Krohn, and T. S. Roy, "On Statistical Aspects of Qjets," *JHEP*, vol. 1501, p. 022, 2015. DOI: 10.1007/JHEP01(2015)022. arXiv: 1409.6785 [hep-ph].

[127] J. Butterworth, B. Cox, and J. R. Forshaw, "WW scattering at the CERN LHC," *Phys.Rev.*, vol. D65, p. 096014, 2002. DOI: 10.1103/PhysRevD.65.096014. arXiv: hep-ph/0201098 [hep-ph].

[128] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets," *JHEP*, vol. 0804, p. 005, 2008. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[129] S. Rappoccio, "A new top jet tagging algorithm for highly boosted top jets," CERN, Geneva, Tech. Rep. CMS-CR-2009-255, Aug. 2009. [Online]. Available: https://cds.cern.ch/record/1358770.

[130] C. Anders, C. Bernaciak, G. Kasieczka, T. Plehn, and T. Schell, "Benchmarking an even better top tagger algorithm," *Phys.Rev.*, vol. D89, no. 7, p. 074047, 2014. DOI: 10.1103/PhysRevD.89.074047. arXiv: 1312.1504 [hep-ph].

[131] T. Plehn and M. Spannowsky, "Top Tagging," *J.Phys.*, vol. G39, p. 083001, 2012. DOI: 10.1088/0954-3899/39/8/083001. arXiv: 1112.4441 [hep-ph].

[132] D. E. Soper and M. Spannowsky, "Finding physics signals with shower deconstruction," *Phys.Rev.*, vol. D84, p. 074002, 2011. DOI: 10.1103/PhysRevD.84.074002. arXiv: 1102.3480 [hep-ph].

[133] D. E. Soper and M. Spannowsky, "Finding physics signals with event deconstruction," *Phys.Rev.*, vol. D89, no. 9, p. 094005, 2014. DOI: 10.1103/PhysRevD.89.094005. arXiv: 1402.1189 [hep-ph].

[134] M. Backovic, J. Juknevich, and G. Perez, "Boosting the Standard Model Higgs Signal with the Template Overlap Method," *JHEP*, vol. 1307, p. 114, 2013. DOI: 10.1007/JHEP07(2013)114. arXiv: 1212.2977 [hep-ph].

[135] D. Krohn, J. Thaler, and L.-T. Wang, "Jets with Variable R," *JHEP*, vol. 0906, p. 059, 2009. DOI: 10.1088/1126-6708/2009/06/059. arXiv: 0903.0392 [hep-ph].

[136] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput.Phys.Commun.*, vol. 178, pp. 852–867, 2008. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[137] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, *et al.*, "Rivet user manual," *Comput.Phys.Commun.*, vol. 184, pp. 2803–2819, 2013. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

[138] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder," *Phys.Lett.*, vol. B641, pp. 57–61, 2006. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210 [hep-ph].

[139] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur.Phys.J.*, vol. C72, p. 1896, 2012. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[140] M. R. Buckley, T. Plehn, and M. Takeuchi, "Buckets of Tops," *JHEP*, vol. 1308, p. 086, 2013. DOI: 10.1007/JHEP08(2013)086. arXiv: 1302.6238 [hep-ph].

[141] M. R. Buckley, T. Plehn, T. Schell, and M. Takeuchi, "Buckets of Higgs and Tops," *JHEP*, vol. 1402, p. 130, 2014. DOI: 10.1007/JHEP02(2014)130. arXiv: 1310.6034 [hep-ph].

[142] R. Barbier, C. Berat, M. Besancon, M. Chemtob, A. Deandrea, *et al.*, "R-parity violating supersymmetry," *Phys.Rept.*, vol. 420, pp. 1–202, 2005. DOI: 10.1016/j.physrep.2005.08.006. arXiv: hep-ph/0406039 [hep-ph].

[143] J. Fan, M. Reece, and J. T. Ruderman, "Stealth Supersymmetry," *JHEP*, vol. 1111, p. 012, 2011. DOI: 10.1007/JHEP11(2011)012. arXiv: 1105.5135 [hep-ph].

[144] S. Gopalakrishna, T. Mandal, S. Mitra, and G. Moreau, "LHC Signatures of Warped-space Vector-like Quarks," *JHEP*, vol. 1408, p. 079, 2014. DOI: 10.1007/JHEP08(2014)079. arXiv: 1306.2656 [hep-ph].

[145] A. Girdhar, B. Mukhopadhyaya, and M. Patra, "Distinguishing Signatures of top-and bottom-type heavy vectorlike quarks at the LHC," *Phys.Rev.*, vol. D91, no. 5, p. 055 015, 2015. DOI: 10.1103/PhysRevD.91.055015. arXiv: 1404.3374 [hep-ph].

[146] S. Yang, J. Jiang, Q.-S. Yan, and X. Zhao, "Hadronic b' search at the LHC with top and W taggers," *JHEP*, vol. 1409, p. 035, 2014. DOI: 10.1007/JHEP09(2014)035. arXiv: 1405.2514 [hep-ph].

[147] V. Khachatryan *et al.*, "Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure," 2015. arXiv: 1503.01952 [hep-ex].

[148] G. Aad *et al.*, "Search for pair and single production of new heavy quarks that decay to a Z boson and a third-generation quark in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *JHEP*, vol. 1411, p. 104, 2014. DOI: 10.1007/JHEP11(2014)104. arXiv: 1409.5500 [hep-ex].

[149] G. Aad *et al.*, "Search for vector-like B quarks in events with one isolated lepton, missing transverse momentum and jets at $\sqrt{s} = 8$ TeV with the ATLAS detector," *Phys.Rev.*, vol. D91, p. 112 011, 2015. DOI: 10.1103/PhysRevD.91.112011. arXiv: 1503.05425 [hep-ex].

[150] T. A. collaboration, "Search for production of vector-like quark pairs and of four top quarks in the lepton plus jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," 2015.

[151] S. Chatrchyan *et al.*, "Search for a Vector-like Quark with Charge 2/3 in $t + Z$ Events from pp Collisions at $\sqrt{s} = 7$ TeV," *Phys.Rev.Lett.*, vol. 107, p. 271 802, 2011. DOI: 10.1103/PhysRevLett.107.271802. arXiv: 1109.4985 [hep-ex].

[152] S. Chatrchyan *et al.*, "Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV," *Phys.Lett.*, vol. B729, pp. 149–171, 2014. DOI: 10.1016/j.physletb.2014.01.006. arXiv: 1311.7667 [hep-ex].

[153] C. Collaboration, "Search for vector-like quarks in final states with a single lepton and jets in pp collisions at sqrt s = 8 TeV," 2014.

[154] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *JHEP*, vol. 1106, p. 128, 2011. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[155] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 0605, p. 026, 2006. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[156] S. Ovyn, X. Rouby, and V. Lemaitre, "DELPHES, a framework for fast simulation of a generic collider experiment," 2009. arXiv: 0903.2225 [hep-ph].

[157] J. de Favereau *et al.*, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," *JHEP*, vol. 1402, p. 057, 2014. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].

[158] N. Kidonakis and R. Vogt, "The Theoretical top quark cross section at the Tevatron and the LHC," *Phys.Rev.*, vol. D78, p. 074 005, 2008. DOI: 10.1103/PhysRevD.78.074005. arXiv: 0805.3844 [hep-ph].

[159] G. Bevilacqua, M. Czakon, C. Papadopoulos, R. Pittau, and M. Worek, "Assault on the NLO Wishlist: pp —¿ t anti-t b anti-b," *JHEP*, vol. 0909, p. 109, 2009. DOI: 10.1088/1126-6708/2009/09/109. arXiv: 0907.4723 [hep-ph].

[160] M. Worek, "NLO mass effects in b anti-b b anti-b production at the LHC," *PoS*, vol. RADCOR2013, p. 038, 2013. arXiv: 1311.2396 [hep-ph].

[161] C. Collaboration, [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMSPublic/MultipleConeSizes14.

[162] A. Katz, M. Son, and B. Tweedie, "Jet Substructure and the Search for Neutral Spin-One Resonances in Electroweak Boson Channels," *JHEP*, vol. 1103, p. 011, 2011. DOI: 10.1007/JHEP03(2011)011. arXiv: 1010.5253 [hep-ph].

[163] A. J. Larkoski, F. Maltoni, and M. Selvaggi, "Tracking down hyper-boosted top quarks," *JHEP*, vol. 1506, p. 032, 2015. DOI: 10.1007/JHEP06(2015)032. arXiv: 1503.03347 [hep-ph].

[164] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, S08003, 2008. DOI: 10.1088/1748-0221/3/08/S08003.

[165] G. Aad *et al.*, "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC," *New J.Phys.*, vol. 13, p. 053 033, 2011. DOI: 10.1088/1367-2630/13/5/053033. arXiv: 1012.5104 [hep-ex].

[166] G. Aad *et al.*, "Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC," *Eur.Phys.J.*, vol. C73, no. 3, p. 2305, 2013. DOI: 10.1140/epjc/s10052-013-2305-1. arXiv: 1203.1302 [hep-ex].

[167] G. Aad *et al.*, "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector," *JHEP*, vol. 1309, p. 076, 2013. DOI: 10.1007/JHEP09(2013)076. arXiv: 1306.4945 [hep-ex].

[168] T. Plehn, G. P. Salam, and M. Spannowsky, "Fat Jets for a Light Higgs," *Phys.Rev.Lett.*, vol. 104, p. 111 801, 2010. DOI: 10.1103/PhysRevLett.104.111801. arXiv: 0910.5472 [hep-ph].

[169] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," *PoS*, vol. ACAT, p. 040, 2007. arXiv: physics/0703039 [PHYSICS].

[170] W. J. Waalewijn, "Calculating the Charge of a Jet," *Phys.Rev.*, vol. D86, p. 094 030, 2012. DOI: 10.1103/PhysRevD.86.094030. arXiv: 1209.3019 [hep-ph].

[171] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, "Calculating Track-Based Observables for the LHC," *Phys.Rev.Lett.*, vol. 111, p. 102 002, 2013. DOI: 10.1103/PhysRevLett.111.102002. arXiv: 1303.6637 [hep-ph].

[172] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, "Calculating Track Thrust with Track Functions," *Phys.Rev.*, vol. D88, p. 034 030, 2013. DOI: 10.1103/PhysRevD.88.034030. arXiv: 1306.6630 [hep-ph].

[173] A. J. Larkoski and J. Thaler, "Unsafe but Calculable: Ratios of Angularities in Perturbative QCD," *JHEP*, vol. 1309, p. 137, 2013. DOI: 10.1007/JHEP09(2013)137. arXiv: 1307.1699 [hep-ph].

[174] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys.Rept.*, vol. 97, pp. 31–145, 1983. DOI: 10.1016/0370-1573(83)90080-7.

[175] B. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference," *Nucl.Phys.*, vol. B238, p. 492, 1984. DOI: 10.1016/0550-3213(84)90333-X.

[176] G. Marchesini and B. Webber, "Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation," *Nucl.Phys.*, vol. B310, p. 461, 1988. DOI: 10.1016/0550-3213(88)90089-2.

[177] J.-C. Winter, F. Krauss, and G. Soff, "A Modified cluster hadronization model," *Eur.Phys.J.*, vol. C36, pp. 381–395, 2004. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085 [hep-ph].

[178] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, *et al.*, "Herwig++ Physics and Manual," *Eur.Phys.J.*, vol. C58, pp. 639–707, 2008. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[179] T. A. collaboration, "A search for $\mathbf{t\bar{t}}$ resonances using lepton plus jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," 2015.

[180] G. Bell, J. Kuhn, and J. Rittinger, "Electroweak Sudakov Logarithms and Real Gauge-Boson Radiation in the TeV Region," *Eur.Phys.J.*, vol. C70, pp. 659–671, 2010. DOI: 10.1140/epjc/s10052-010-1489-x. arXiv: 1004.4117 [hep-ph].

[181] J. R. Christiansen and T. Sjostrand, "Weak Gauge Boson Radiation in Parton Showers," *JHEP*, vol. 1404, p. 115, 2014. DOI: 10.1007/JHEP04(2014)115. arXiv: 1401.5238 [hep-ph].

[182] F. Krauss, P. Petrov, M. Schoenherr, and M. Spannowsky, "Measuring collinear W emissions inside jets," *Phys.Rev.*, vol. D89, no. 11, p. 114 006, 2014. DOI: 10.1103/PhysRevD.89.114006. arXiv: 1403.4788 [hep-ph].

[183] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, "Jet Charge at the LHC," *Phys.Rev.Lett.*, vol. 110, no. 21, p. 212 001, 2013. DOI: 10.1103/PhysRevLett.110.212001. arXiv: 1209.2421 [hep-ph].

[184] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders," *Z.Phys.*, vol. C45, p. 109, 1989. DOI: 10.1007/BF01552335,10.1007/BF01556677.

[185] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, *et al.*, "Theory and phenomenology of two-Higgs-doublet models," *Phys.Rept.*, vol. 516, pp. 1–102, 2012. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[186] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, *et al.*, "Event generation with SHERPA 1.1," *JHEP*, vol. 0902, p. 007, 2009. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

[187] T. Gleisberg and S. Hoeche, "Comix, a new matrix element generator," *JHEP*, vol. 0812, p. 039, 2008. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[188] G. Isidori, A. V. Manohar, and M. Trott, "Probing the nature of the Higgs-like Boson via $h \to V\mathcal{F}$ decays," *Phys. Lett.*, vol. B728, pp. 131–135, 2014. DOI: 10.1016/j.physletb.2013.11.054. arXiv: 1305.0663 [hep-ph].

Self-References

[A] M. Stoll, "Vetoed jet clustering: The mass-jump algorithm," *JHEP*, vol. 1504, p. 111, 2015. DOI: 10.1007/JHEP04(2015)111. arXiv: 1410.4637 [hep-ph].

[B] K. Hamaguchi, S. P. Liew, and M. Stoll, "Jumping into buckets, or How to decontaminate overlapping fat jets," *Phys. Rev.*, vol. D92, no. 1, p. 015 012, 2015. DOI: 10.1103/PhysRevD.92.015012. arXiv: 1505.02930 [hep-ph].

[C] M. Spannowsky and M. Stoll, "Tracking New Physics at the LHC and beyond," 2015. arXiv: 1505.01921 [hep-ph].

[D] M. Endo, K. Hamaguchi, K. Ishikawa, and M. Stoll, "Reconstruction of Vector-like Top Partner from Fully Hadronic Final States," *Phys.Rev.*, vol. D90, no. 5, p. 055 027, 2014. DOI: 10.1103/PhysRevD.90.055027. arXiv: 1405.2677 [hep-ph].

[E] *FastJet Contrib*, http://fastjet.hepforge.org/contrib/.

[F] *HPTTopTagger, HPTWTagger, and HPTZTagger*, https://www.ippp.dur.ac.uk/~mspannow/webippp/HPTTaggers.html.

Acknoledgments

First and foremost I would like to thank my supervisor Koichi Hamaguchi for accepting me as his PhD student and anabling me to enter the University of Tokyo. I am grateful for his constant support on physics and private matters. In particular his insightful comments and guiding questions were invaluable.
I highly appreciate the help from my collaborators during various research projects, namely Motoi Endo, Kazuya Ishikawa, Seng-Pei Liew, and Michael Spannoswky, from whom I could learn a lot.
Also I want to thank Michihisa Takeuchi for fruitful discussions. I thank Michael Spannoswky for inviting me to Durham, and Tilman Plehn for his invitation to Heidelberg.
All the members of the High Energy Physics Theory Group have created a relaxed and enjoyable atmosphere, for which I am grateful.
I also thank the "Leading Graduate Course for Frontiers of Mathematical Sciences and Physics" (FMSP) for their financial and immaterial support during the PhD course, and enabling me to attend a number of conferences.
Last but not least, I thank my parents and relatives for supporting and believing in me all those years. I thank my friends for many good memories and forgivingness for not being able to meet and have a good time more often.
Finally I am more than grateful to Yuka. She made these years in Tokyo really special.

Tokyo, August 2015
Martin Stoll