論文の内容の要旨

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）

Name:

シュトル　マルティン

Stoll, Martin

One of the main purposes of the Large Hadron Collider (LHC) is to investigate the mechanism behind electroweak symmetry breaking. Its first run has been a huge success with the discovery of a Higgs boson that so far is compatible with the Standard Model of Particle Physics (SM). Although the SM is renormalizable and may be valid up to the Planck scale, delicate fine tuning is necessary to cancel the large loop corrections to the Higgs mass. One way in which this problem can be solved is to include additional heavy particles that cut off these loop contributions. In this spirit many models beyond the SM (collectively called BSM) introduce new heavy resonances, in particular top partners as the largest loop contribution stems from the top quark. Characteristic signatures of these new resonance decays include top quarks, the Higgs boson, or electroweak gauge bosons. Efficient tagging, i.e. the identification of these SM resonances with simultaneous rejection of (QCD) backgrounds, is hence crucial for BSM searches. But also for Higgs boson analyses within the SM, the top quark plays an important role, as it is the heaviest of all elementary particles and thus possesses the largest coupling to the Higgs boson.

More specifically, the new resonances introduced to reduce fine tuning are expected to have TeV-scale masses. Their decay products (like top and Higgs) are therefore boosted. The term “boost” was coined for particles whose transverse momentum exceeds twice their rest mass, $p_\perp \gtrsim 2m$. Phenomenology significantly changes with boost, as the subsequent decay products are now collimated in the detector. Essentially every current top tagging algorithm aims to collect all this radiation in one large-radius "fat" jet and analyze its substructure. A prime goal of the upcoming run II at the LHC is the search for heavy new resonances, and taggers developed for boosted tops, Higgses, etc. are expected to play a major role in these analyses. While recent progress has been impressive, not quite so much focus has yet been put on tagging SM resonances at the very energy frontier to be accessible 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 take a different approach. It generalizes the so-called "mass-drop" subjet identification inside a fat jet, making it applicable to scenarios where the construction of fat jets is troublesome. The mass-drop algorithm starts from a fat jet that has been built with a sequential jet algorithm, i.e. through a sequence of $2 \to 1$ intermediate protojet mergings. Then the fat jet is unclustered again, and at each step $j \to j_1 j_2$ the condition $\max(m_{j_1}, m_{j_2}) < \theta \cdot m_j$ is checked. If it is fulfilled and a mass drop occurred, both prongs are expected to belong to the decay products of a heavy resonance. Otherwise the less massive one is discarded as noise. Prongs with invariant mass below a parameter μ are not further unclustered and constitute the set of subjets.

Our new approach is bottom-up and does not require a fat jet. It starts from the final-state particles (calorimeter towers or particle flow in the experiments) and applies sequential jet clustering $ij \to k$ as long as the new protojet is still light, $m_k < \mu$. Otherwise a "mass-jump" veto is tested, which reads $\theta \cdot m_k > \max(m_i, m_j)$. If it is fulfilled, the two prongs are identified to be the decay products of a heavy resonance, which may be identified with k. These vetoed prongs do not participate further in the clustering process. The top quark requires one more ingredient, because its hadronical decay yields three distinct subjets. Hence we introduce the possibility of a veto between an "active" protojet and one that has already been vetoed. Our algorithm produces jets with variable radii depending on the jet's vicinity, which is a novelty.

We examine the mass-jump algorithm and find many favourable qualities. Most straightforwardly, it can replace the mass-drop step in the popular `HEPTopTagger`, and results can easily be compared between the original and our modified tagger. We investigate the efficiency of top quark tagging versus QCD background rejection, and find that the working point with largest top tagging efficiency gives identical results. When turning to high-purity working points, however, our modified algorithm including the mass-jump algorithm yield superior results. About the origin of this improvement we speculate that it might have to do with the cascade decay of the top quark, $t \to bW^+ \to bjj$. When the corresponding fat jet is declustered, the successive mass drops involve very different scales: $m_t/m_W \sim 0.5$ and $m_W/m_j \sim 0$. Hence the threshold θ has to be chosen with both scales in mind. The mass-jump algorithm, on the other hand, does not face this problem, because all "jumps" are between light protojets. We verify that in the relevant parameter regions (θ,μ), well-isolated jets are not affected much by the veto.

Turning to the energy frontier, this weakness of fat jet unclustering algorithms can pose severe problems. We investigate the pair production of TeV-scale vectorlike tops as predicted for example by Little Higgs models. If each vectorlike top decays into a top quark and a Higgs boson, the event contains four boosted electroweak-scale resonances. They are not well separated any more, affecting fat-jet-based analyses in several ways. First, unique and isolated fat jets cannot be built. If their size is chosen too small, not all decay products are included and taggers will fail. If they are chosen large enough, they typically contain additional hard radiation from other resonances. Subjet finding then may be aggravated. These problems are avoided by the mass-jump algorithm. It identifies the individual jets corresponding to final-state light quarks without defining intermediate fat jets. All mass jumps involve similar scales, and the reconstruction of isolated jets is not impeded by the veto. We focus on kinematic reconstruction of the complete process and show that this is possible with good quality during the first stage of run II at the LHC even for vectorlike top masses above 1 TeV. If similar particles do show up at the LHC, our novel algorithm can increase the discovery reach and the accuracy of a measurement. We note that the assignment of at least ten jets to their respective top quarks and Higgs bosons poses a combinatorial problem.

Our suggestion is to collect the jets for each candidate in a "bucket" and use a decoupled, χ^2-like metric based on the buckets' invariant masses. Despite being a simple idea, the results show very good performance. In particular our analysis outperforms its counterpart based on a common jet clustering algorithm (without mass jump) and a similar analysis based on fat jets.

Large final state multiplicity is not the only facet of the energy frontier to be accessible soon. The high collision energy allows to produce very heavy neutral gauge bosons Z' or charged W' with a mass of several TeV. When decaying into a pair of electroweak-scale particles, these are highly boosted and thus very much collimated in the detector. Resonances with a very large transverse momentum force us to reconsider building jets from calorimetric 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 $p_\perp \sim 1$ TeV or electroweak gauge bosons above $p_\perp \sim 500$ GeV. In this thesis we also develop dedicated tagging algorithms based on fat jets is the highly boosted regime. Fat jet contamination does not pose a problem for such high boosts. We focus on top quarks as well as W and Z bosons. The essential ingredient is to combine calorimetric information with the good spatial resolution of the charged particle tracker. Basing a tagger solely on tracks, however, is not efficient. The observed tracks lack the information of all neutral hadrons and leptons and do not give a good measure of the energy deposit in the fat jet. The calorimetric information, on the other hand, does, and we start from a local re-calibration of the tracks using the calorimeter-based jet. Our algorithm differs from state-of-the-art top taggers (designed for low and intermediate boosts) in several key aspects. First of all, we do not decluster the (track-based) fat jet as is customary in those taggers, but recluster the subjets directly with a characteristic radius. The first reason is that the high collimation allows for smaller-size fat jets that do not suffer much from typical noise. Furthermore, none of the subjets can become very soft as may be the case for mildly-boosted tops, leaving a clearer substructure. Also we do not explicitly search for top-like substructure, but rely on the three (two) hardest subjets for top quarks (W and Z bosons) to reduce shaping of the background. These facets clearly distinguish our algorithms from their spiritual predecessors.

We find good tagging efficiencies and background rejection power for all algorithms; in addition we find the reliable reconstruction of the resonance mass. We also investigate possible discovery scenarios of very heavy new resonances at the LHC and a future 100 TeV collider based on our tagging algorithms. The first benchmark scenario is a heavy Z' dominantly decaying into a pair of top quarks, as predicted by topcolour models. At the LHC, it can be excluded up to masses above 3 TeV with the use of our `HPTTopTagger`. Production of a heavy charged boson and its decay $W'^{\pm} \to W^{\pm}Z$ constitutes the second benchmark scenario. It is analyzed using our `HPTWTagger` and `HPTZTagger`, depending on which of the two highly boosted SM gauge bosons decays hadronically. We find that the hypothetical W' can be excluded up to masses above 4 TeV. These discovery reaches are significantly larger than current exclusion bounds and similar analyses are difficult with common tagging algorithms that do not make use of charged tracks. Note that the highly boosted regime does not necessarily involve very heavy resonances: The rare prompt decay of the Higgs boson into a Z boson and a CP-odd scalar A yields a highly boosted A if it is very light. Those particles are predicted by models that contain more than one Higgs doublet. To demonstrate this we analyze this very process with the help of charged tracks and predict strong bounds.

In summary, we develop a new class of jet clustering algorithms, the mass-jump algorithm, that has beneficial properties in busy environments. In a realistic process we show that our analysis based on mass jump outperforms common approaches. With this new tool the discovery reach of the LHC can be extended on the high-multiplicity energy frontier. Extending the discovery reach on the other energy frontier as well is achieved by our tagging algorithms in the highly boosted regime. Both regimes are crucial for the success of the LHC. To investigate the mechanism of electroweak symmetry breaking and the Higgs sector it is necessary to restrict the plethora of models beyond the SM. Direct searches for new resonances in the TeV regime are a powerful means to do so. Tagging algorithms for boosted top quarks, Higgs and electroweak gauge bosons will play a key role.

We finally note that all algorithms developed in this thesis have been made publicly available.