IL NUOVO CIMENTO Vol. 33 C, N. 4 Luglio-Agosto 2010
DOI 10.1393/ncc/i2010-10679-8

Colloquia: TOP2010

Boosted tops: Physics and reconstruction

E. Chabert on behalf of the CMS Collaboration(1) and J. Sjölin on behalf of the ATLAS Collaboration(2)

(1) *Interuniversity Institute for High Energies, Université Libre de Bruxelles Bruxelles, Belgium*
(2) *University of Stockholm - Stockholm, Sweden*

(ricevuto il 18 Agosto 2010; approvato il 18 Agosto 2010; pubblicato online il 26 Ottobre 2010)

Summary. — The production of boosted top quarks in the high centre-of-mass energy collisions at the LHC is a probe to new phenomena at the TeV scale. Numerous extensions to the Standard Model (SM) predict massive particles whose couplings to top quarks are enhanced. Such massive particles, referred as $t\bar{t}$ resonances, will distort the $t\bar{t}$ invariant mass spectrum relative to the SM expectation. This presentation is dedicated to studies performed by the ATLAS and CMS Collaborations. New techniques were developed to reconstruct boosted top quarks as single objects, both as a top-tagged jet with a substructure in the hadronic decay mode and a jet with a lepton inside in the leptonic decay mode. These techniques were applied in dedicated analyses to search for high mass $t\bar{t}$ resonances. Expected limits and the potential for observation are presented for a luminosity of few hundreds of pb^{-1}.

PACS 14.65.Ha – Top quarks.

1. – Physics motivations

The high luminosity and centre-of-mass energy of the proton-proton collisions at the LHC will produce a large number of top quark pairs. These pairs which are produced via Standard Model (SM) strong interaction processes, gluon fusion or quark/anti-quark annihilation, are expected to be the main source of boosted top quarks, *i.e.* top quarks with a large boost in the laboratory system. However, the production of high-energy top quarks is also a generic signature of many models for physics beyond the Standard Model (BSM). It is therefore important to prepare for searches of resonances decaying into $t\bar{t}$ to probe possible scenarios of new physics with the ATLAS [1] and CMS [2] detectors.

1·1. *Production of boosted top quarks.* – Due to its large mass, of the order of the electro-weak symmetry breaking (EWSB) scale, the top quark plays a special role in many EWSB BSM theories. Numerous extensions to the SM predict gauge interactions whose couplings with the third-generation quarks, and in particular the top quark, are enhanced. In models with top condensation such as technicolor and topcolor, the role of the SM

Higgs boson is filled by a $\mathrm{t\bar{t}}$ bound state [3]. These models predict additional heavy gauge bosons, which couple strongly to top quarks like the colour-singlet Z′ [4]. In models with extra dimensions, such as Rundall-Sundrum [5] and ADD models [6], TeV-scale gravitons can decay, in some cases preferentially, to top pairs [7]. In all these cases, the production of top pairs at hadron colliders through BSM mechanisms distorts the $\mathrm{t\bar{t}}$ invariant mass ($M_{\mathrm{t\bar{t}}}$) spectrum relative to the SM expectation [7]. Signals of new physics in top pair production have already been searched for at the Tevatron experiments [8,9]. Mass limits for a narrow topcolor leptophobic resonance range from 720 to 800 GeV.

1˙2. *Topology of boosted top quarks*. – The experimental signature of top quarks depends on their boost in the laboratory frame. The dominant production of top quark pairs via strong interaction produce mainly low-p_{T} top quarks. In that regime, the decay products can be clearly identified with standard reconstruction techniques. The experimental signature of a leptonic decay of a top quark is an isolated lepton, transverse missing energy and a jet which can be tagged by b-tagging algorithms based on the presence of a secondary vertex or the impact parameter of displaced tracks associated to the jet. The experimental signature of a hadronic decay is based on three jets with a p_{T} higher than 30–40 GeV, one being potentially b-tagged. In the case of high-p_{T} top quarks, these requirements would fail due to the boosted decay products. The angles between the decay products and the initial top quark tend to be smaller when the top quark is boosted, rendering classical top identification methods inefficient. The standard isolation criteria for leptons cannot be used anymore because the lepton can be contained inside the jet cone associated to the bottom quark. On the hadronic side, depending on the boost, the jets could overlap, or even collapse into a single jet. These topological considerations guided the development of new strategies to probe the production of boosted top quarks with dedicated reconstruction techniques as described in the next section.

2. – Reconstruction of boosted top quarks

2˙1. *Hadronic decay mode*. – The general strategy for tagging boosted top quarks decaying hadronically is to identify jet substructures in top-quark jets, and to use these substructures to impose kinematic cuts that discriminate against non-top jets of the same p_{T}.

The current section describes the strategy set up in the CMS Collaboration [10] to tag boosted top quarks following the prescriptions of [11]. It takes advantage of the high granularity of the electromagnetic and hadron calorimeters to identify jet substructures. Several sequential recombination algorithms present in the FASTJET package [12] are tried: k_{T} [13], anti-k_{T} [14], and the Cambridge-Aachen (C-A) algorithm [15]. While the (anti-)k_{T} algorithm preferentially merges constituents with (high) low transverse momentum with respect to their nearest neighbors, the C-A algorithm relies only on distances. The R parameter is taken to be 0.8 for all three algorithms, which is sufficiently large to include all the towers from a top cascade decay if the top momentum exceeds 800 GeV. The C-A algorithm selects the subjets closest to the hard jet axis compared to the other algorithms and is therefore well suited to discriminate softer subjets within harder jets [11].

Hadronic top tagging algorithm. The C-A jets are required to have $p_{\mathrm{T}} > 250$ GeV, and rapidity $|\eta| < 2.5$. Then a sequential declustering procedure in two steps is applied, leading to up to four subjets. During the procedure, soft clusters are ignored with the

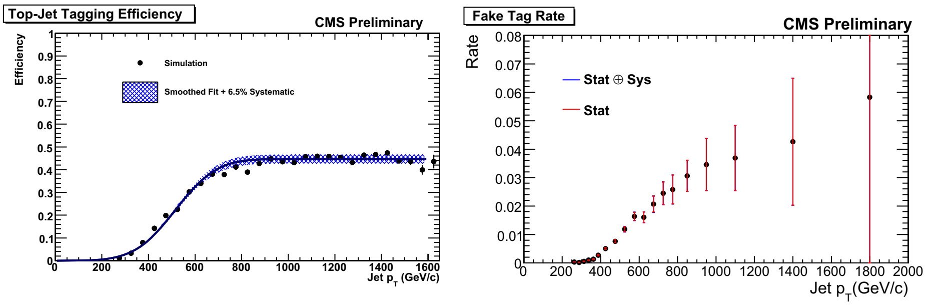

Fig. 1. – Efficiency for matched top-jets on the left as a function of jet p_{T}, including the 6.5% systematic uncertainty. The efficiency is fit to a functional form of $0.45 \times \frac{1}{2}(\mathrm{Erf}(\frac{p_{\mathrm{T}}-516}{197}))$. Fake top-tag parametrization *vs.* jet p_{T} for a $100\,\mathrm{pb}^{-1}$ scenario on the right. The central value is taken as the prediction from simulation, and the uncertainties are given as the expected statistical plus systematic uncertainties.

use of a criterion $p_{\mathrm{T}}^{\mathrm{cluster}} > 0.05 \times p_{\mathrm{T}}^{\mathrm{hard\,jet}}$. A successfully decomposed jet yields either three or four clusters, the other jets are no longer considered.

As a hadronic top decay consists of a bottom quark and two additional quarks coming from the W decay it is natural to impose kinematic cuts to the jet and its substructure. The mass of the jet is required to be consistent with the top mass: $100\,\mathrm{GeV} < m_{\mathrm{jet}} < 250\,\mathrm{GeV}$, discriminating the generic QCD non-top jets where the mass scales with the jet p_{T}. The three highest p_{T} subjets are taken pairwise, and the minimum invariant mass of those six pairwise candidates (m_{min}) is required to be higher than 50 GeV, discriminating the jets where no on-shell W is produced. The discrimination with the use of C-A algorithm is superior to the one obtained with other algorithms: $S/\sqrt{B} = 2.4$ for C-A, 1.6 for k_{T}, and 1.3 for anti-k_{T} for a luminosity of $100\,\mathrm{pb}^{-1}$, the signal being matched top-jets coming from $\mathrm{Z}' \to \mathrm{t}\bar{\mathrm{t}}$ with $M = 2\,\mathrm{TeV}/c^2$ and the background being jets coming from generic QCD muti-jet processes.

Hadronic algorithm performance. Non-top decays may pass the previously defined selection and thus fake a boosted top quark. In order to derive a parametrization of the fake tag rate, a data-driven method is proposed that makes use of a high-statistics sample, and uses an "anti-tag and probe" method. This method is expected to provide over a thousand fake tags for a data sample of $100\,\mathrm{pb}^{-1}$, allowing for a robust data-driven determination of this background. The right plot of fig. 1 shows the number of events and the fake tag parametrization as a function of jet p_{T} for a $100\,\mathrm{pb}^{-1}$ data sample. With this luminosity, it is possible to reliably estimate the fake tag rate directly from the data, with an approximately 33% statistical uncertainty for jets with $p_{\mathrm{T}} = 800\,\mathrm{GeV}$.

The efficiency of the boosted top algorithm is difficult to compute from data. It has been estimated that a luminosity of around $12.5\,\mathrm{fb}^{-1}$ would be needed to estimate the efficiency from data with a statistical uncertainty of around 10%, based on the continuum semileptonic $\mathrm{t}\bar{\mathrm{t}}$. Instead, the efficiency for early data is estimated from simulation. Several systematic effects can affect the estimate of the tagging efficiency due to uncertainties in the shower development which influences the profile of the subjets. The total uncertainty of the effects of initial and final state radiation, renormalization scale, and fragmentation is estimated to be 3.8%. Additionally, detector-based systematic uncertainties including jet energy and angular resolutions are conservatively estimated

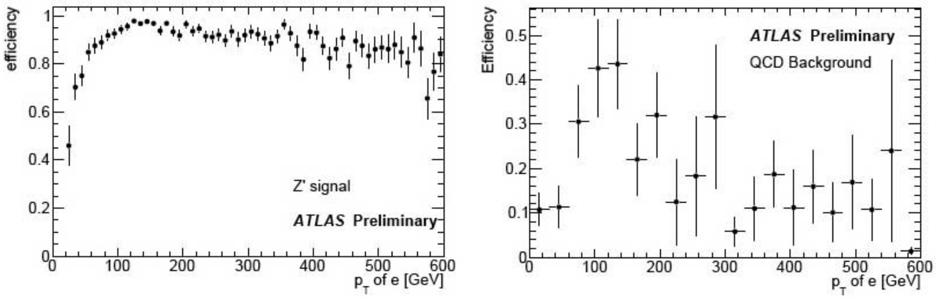

Fig. 2. – Electron top tag efficiency for signal and QCD as a function of the reconstructed electron p_{T}.

to be 5.3%, The left plot of fig. 1 shows the efficiency with statistical uncertainties, as well as the total 6.5% systematic uncertainty, obtained from combining the theoretical (3.8%) and detector-based (5.3%) systematic uncertainties.

2˙2. *Leptonic decay mode*. – We also tag boosted top jets where the top decays to leptons and a b-jet. The mass relation to the top mass using the decay products is lost due to the escaping neutrino, but one can still form a variable related to fraction of the visible mass carried by the muon [16] using $x_l = 1 - m_b/m_{\text{vis}}^2$, where m_{vis} is the mass of the b-jet and the lepton. For leptonic top decays this variable is peaked towards 1. Another example of observable that has discriminating power against non-top jets is $y_b = p_{l\perp b} \times \Delta R(l,b)$, where $p_{l\perp b}$ is the p_{T} with respect to the b-jet and $\Delta R(l,b)$ is the opening angle (in η, ϕ space). It is similar to the square root of the $k_\perp$ distance and is expected to have small values for QCD jets. Both muon and electron decays can be used if the electro-magnetic calorimeter has fine enough granularity.

This section describes a leptonic top tagging algorithm studied in ATLAS [17]. It uses exactly the leptonic observables explained in the previous paragraph.

Leptonic top tagging algorithm and performance. The algorithm starts by selecting the hardest electron or muon in the event with $p_{\mathrm{T}} > 20\,\text{GeV}$ within $|\eta| < 2.5$. Then the $k_\perp$ algorithm, calculated with the FastJet [12] package, with the parameter $D = 0.6$ is run on topological clusters from the calorimeters. The resulting jet within $\Delta R(l,b) < 0.6$ is selected if it has $p_{\mathrm{T}} > 300(200)\,\text{GeV}$ for electron (muon) events. The first splitting scale of the jet defined as the squared root of the minimal $k_\perp$ distance corresponding to the recombination from 2 to 1 subjets is required to be at least 10 GeV. Cuts are then applied to x_l and y_l to reduce fake tags from QCD. The W mass constraint is inferred to solve for the neutrino longitudinal-component (p_z^ν) followed by requiring the solution with smallest real part of p_z^ν and $\Delta R(\nu_l, l) < 1$, and as the last step the neutrino is added to the leptonic top 4-momentum. The performance of this first preliminary version of the ATLAS leptonic top tagger is shown in fig. 2 for the electron case.

3. – Search for high mass $\mathrm{t\bar{t}}$ resonances

Here we present studies from both the ATLAS and CMS experiments where we search for high mass $\mathrm{t\bar{t}}$ resonances where we expect a large fraction of the events to contain boosted tops. As a benchmark scenario we use Z' decays to $\mathrm{t\bar{t}}$.

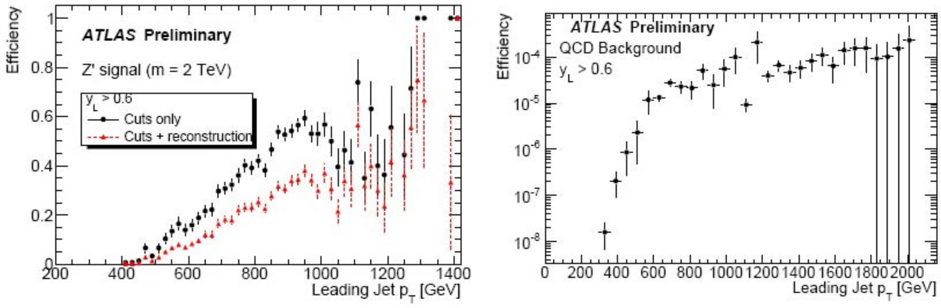

Fig. 3. – Total selection efficiency for a 2 TeV Z' and QCD as a function of the leading jet p_{T}.

3˙1. *Search in the leptonic channel by the ATLAS Collaboration.* – ATLAS has performed a preliminary sensitivity study for narrow high mass $\mathrm{t\bar{t}}$ resonances [17]. This analysis uses the leptonic top tagger described in subsect. **2**˙2 and a hadronic top tagger based on $k_\perp$ jets with $D = 0.6$ and at least a $p_{\mathrm{T}} > 300\,\mathrm{GeV}$. The hadronic tagger uses jet mass, splittings scales from 1 to 2 jets, 2 to 3 jets, and 3 to 4 jets as observables. The observables are combined using likelihood ratios of signal and background probability distributions. The Z' mass is simply reconstructed by forming the invariant mass of the leading leptonic and the hadronic top tagged jets. An example of the performance for a 2 TeV Z' is shown in fig. 3. The Gaussian mass resolution around the peak is 100 GeV. The expected 95% CL limits on the production cross-sections using $1\,\mathrm{fb}^{-1}$ of integrated luminosity at 14 TeV centre-of-mass energy are 0.55 pb for a 2 TeV Z', and 0.16 pb for a 3 TeV Z' using Bayesian statistics and 15% uncertainty on the acceptance and 10% uncertainty on the luminosity.

3˙2. *Search in the leptonic channel by the CMS Collaboration.* – This section presents a search for heavy narrow $\mathrm{t\bar{t}}$ resonances in the muon-plus-jets final states with the CMS detector [18]. The branching ratio of this muon+jets final state is about 15%.

The event selection is based on a non-isolated 15 GeV single-muon trigger and relies on the presence of at least one muon candidate and at least two jets. The muon is required to have a p_{T} larger than 25 GeV, $|\eta| < 2.1$ and a cut on the transverse impact parameter to select prompt muons. As the standard isolation criteria fail in the topologies of highly boosted top quarks, new more efficient cuts have been derived. The new cuts are based on the transverse momentum of the muon relative to the direction of the closest jet and their distance in the (η, ϕ) coordinates.

The calibrated calorimeter jets are reconstructed with the SISCone algorithm [19] (with $R = 0.5$). The leading jet should have a p_{T} exceeding 260 GeV due to the merging of the three quark jets from the hadronic top decay, while the second one should be larger than 50 GeV, both with $|\eta| < 2.5$. For further background suppression, the leptonic transverse energy $H_{\mathrm{T}}^{\mathrm{lep}}$, defined as the scalar sum of transverse energies of the leading muon and $E_{\mathrm{T}}^{\mathrm{miss}}$, is required to exceed 200 GeV.

The search for $\mathrm{t\bar{t}}$ resonances in the invariant mass spectrum employs a full reconstruction of the event. The neutrino transverse momentum is given by the measured missing transverse energy $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, and the longitudinal momentum is determined by imposing the W mass to the system composed by the muon and the neutrino candidate. All possible jet combinations are considered and the full event is reconstructed by minimizing

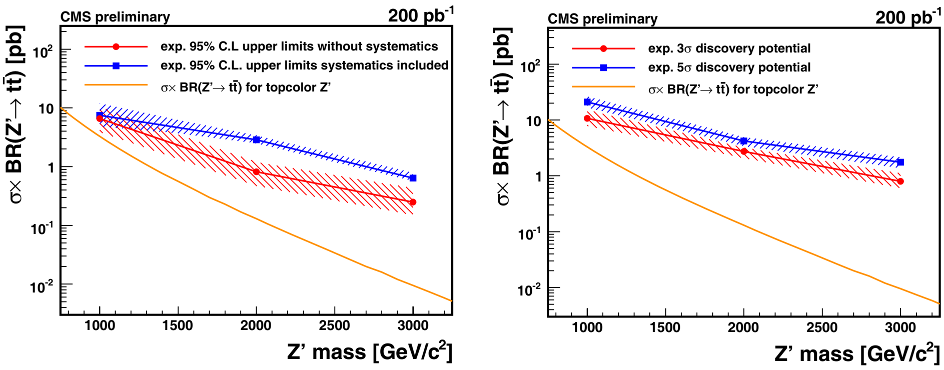

Fig. 4. – Expected limits at 95% CL on the left and 3σ (5σ) observation (discovery) on the right (together with the 1σ confidence interval for these limits) on $\sigma_{Z'} \times \mathrm{BR}(Z' \to t\bar{t})$ obtained by a CMS analysis in the muon+jets channel. An integrated luminosity of $200\,\mathrm{pb}^{-1}$ is assumed to be collected with the CMS detector at $\sqrt{s} = 10\,\mathrm{TeV}$. The cross section of the topcolor Z′ [20] is superimposed.

a global ΔR function. The function takes into account the small separation of the decay products of the semileptonically decaying top quark, the large separation of the two tops in a resonant decay and a term to reduce the tail towards low values of the reconstructed top pair mass.

The dominant background processes for this channel are the continuum $t\bar{t}$ and the production of vector bosons associated with extra jets (W/Z+jets) and the generic multi-jet QCD processes. These background processes are estimated from data using a simultaneous fit on the $M_{t\bar{t}}$ and $H_{\mathrm{T}}^{\mathrm{lep}}$ distributions in the signal region and in a sideband region which is QCD enriched. The effect of different sources of systematic uncertainties are studied. For the theoretical part, the modelling of $t\bar{t}$, W+jets and Z+jets processes and their overall cross-section are considered. For the experimental part, the dominant effect is the knowledge of the jet energy scale which is assumed to be about 10%.

The results presented in fig. 4 include the expected 95% CL upper limits with and without including all sources of uncertainties as a function of $M_{Z'}$ and the lower values of $\sigma_{Z'} \times \mathrm{BR}(Z' \to t\bar{t})$, for which an observation of $Z' \to t\bar{t}$ production is expected with at least 3σ and 5σ, respectively.

3˙3. *Search in the hadronic channel performed by the CMS Collaboration.* – This section presents a search for high-mass resonances ($\geq 1\,\mathrm{TeV}$) decaying into $t\bar{t}$ in the all-hadronic decay mode with the CMS detector. This analysis [21] takes advantage of the high branching ratio which is about 46% and uses the newly developed top-tagging algorithm described in subsect. **2**˙1.

The selection of the events requires two top-tagged jets with a $p_{\mathrm{T}} > 250\,\mathrm{GeV}$, $|\eta| < 2.5$ and the top-tagging selection described in **2**˙1. The background contributions to this analysis are numerically dominated by generic QCD dijets events with two fake tags and then by the continuum $t\bar{t}$. The background uncertainties, estimated to be 36% on the observability of a 2 TeV resonance, are dominated by the jet energy scale uncertainty estimated to be 35% by removing the jet calibration.

The efficiency for heavy resonances is taken from simulation using samples of $Z' \to t\bar{t}$. The systematic uncertainty for a single top-tagged jet is estimated to 6.5%

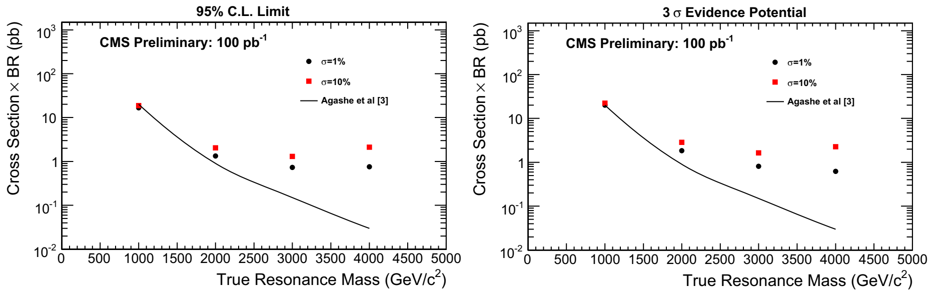

Fig. 5. – Expected limits at 95% C.L. on the left and 3σ observation on the right on $\sigma_{Z'}\times \mathrm{BR}(Z' \to t\bar{t})$ obtained by a CMS analysis in the dijets channel using top-tagged jets. An integrated luminosity of $100\,\mathrm{pb}^{-1}$ is assumed to be collected with the CMS detector at $\sqrt{s} = 10\,\mathrm{TeV}$. The cross section for a Kaluza-Klein gluon for the RS1 model [22] is superimposed as calculated in ref. [23].

(cf. subsect. **2**˙1) and is doubled for a dijet requirement, *i.e.* 13%. To perform the hypothesis test, the amount of signal is estimated in a given mass window accounting for the resolution of the boosted top jets estimated from the simulation of resonances decaying to $t\bar{t}$. The jet energy scale uncertainty on the acceptance is estimated to be 5% and leads together with the uncertainty on the top tagging efficiency to a combined uncertainty of 14%.

The left plot in fig. 5 shows the 95% CL exclusion curve for $\int \mathcal{L}\mathrm{d}t = 100\,\mathrm{pb}^{-1}$. The right plot in fig. 5 shows the 3σ discovery curves for $\int \mathcal{L}\mathrm{d}t = 100\,\mathrm{pb}^{-1}$.

REFERENCES

[1] Aad G. *et al.*, *JINST*, **3** (2008) S08003.
[2] Chatrchyan S. *et al.*, *J. Instrum.*, **3** (2008) S08004.
[3] Cvetic G., *Rev. Mod. Phys.*, **71** (1999) 513.
[4] Rosner Jonathan L., *Phys. Lett. B*, **387** (1996) 113.
[5] Randall Lisa and Sundrum Raman, *Phys. Rev. Lett.*, **83** (1999) 3370.
[6] Arkani-Hamed Nima, Dimopoulos Savas and Dvali G. R., *Phys. Lett. B*, **429** (1998) 263.
[7] Frederix Rikkert and Maltoni Fabio, *JHEP*, **01** (2009) 047.
[8] Aaltonen T. *et al.* (The CDF Collaboration), *Phys. Rev. Lett.*, **100** (2008) 231801.
[9] D0 Collaboration, D0 Note 5882-CONF (2009).
[10] CMS Collaboration, *A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging*, *CMS PAS*, JME-09-001 (2009).
[11] Kaplan D., Rehermann K., Schwartz M. and Tweedie B., *Phys. Rev. Lett*, **101** (2008) 142001.
[12] Salam G., Cacciari M. and Soyez G., *Phys. Lett. B*, **641** (2006) 57.
[13] Cataniand S., Dokshitzerand Y. L., Seymour M. H. and Webber B. R., *Nucl. Phys. B*, **406** (1993) 187.
[14] Cacciari M., Salam G. and Soyez G., *JHEP*, **0804** (2008) 063.
[15] Dokshitzer Y. L., Leder G.D., Moretti S. and Webber B. R., *JHEP*, **9708** (1997) 001.
[16] Thaler J. and Wang L.-T., *JHEP*, **07** (2008) 092.
[17] ATLAS Collaboration, *Reconstruction of high mass $t\bar{t}$ resonances in the lepton+jets channel*, *CERN CDS*, ATL-PHYS-PUB-2009-081 (2009).

[18] CMS Collaboration, *Search for heavy narrow* $\mathrm{t}\bar{\mathrm{t}}$ *resonances in muon-plus-jets final states with the CMS detector*, *CMS PAS*, EXO-09-008 (2009).
[19] Salam Gavin P. and Soyez Gregory, *JHEP*, **05** (2007) 086.
[20] Harris Robert M., Hill Christopher T. and Parke Stephen J., *Cross Section for Topcolor* Z'_t *decaying to* $t\bar{t}$, arXiv:hep-ph/9911288v1 (1999).
[21] CMS Collaboration, *Search for High-Mass Resonances Decaying into Top-Antitop Pairs in the All-Hadronic Mode*, *CMS PAS*, EXO-09-002 (2009).
[22] Randall Lisa and Sundrum Raman, *Phys. Rev. Lett.*, **83** (1999) 3370.
[23] Agashe Kaustubh, Belyaev Alexander, Krupovnickas Tadas, Perez Gilad and Virzi Joseph, *Phys. Rev. D*, **77** (2008) 015003.