

# Performance of Tracking, b-tagging and Jet/MET reconstruction at the CMS High Level Trigger

Mia Tosi

Dipartimento di Fisica ed Astronomia, Università degli Studi di Padova and INFN,  
via Marzolo 8, 35131 Padova (Pd), Italy

E-mail: mia.tosi@gmail.com

**Abstract.** The trigger systems of the LHC detectors play a crucial role in determining the physics capabilities of experiments. In 2015, the center-of-mass energy of proton-proton collisions will reach 13 TeV up to an unprecedented luminosity of  $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . A reduction of several orders of magnitude of the event rate is needed to reach values compatible with detector readout, offline storage and analysis capabilities. The CMS experiment has been designed with a two-level trigger system: the Level-1 Trigger (L1T), implemented on custom-designed electronics, and the High Level Trigger (HLT), a streamlined version of the offline reconstruction software running on a computer farm. A software trigger system requires a trade-off between the complexity of the algorithms, the sustainable output rate, and the selection efficiency. With the computing power available during the 2012 data taking the maximum reconstruction time at HLT was about 200 ms per event, at the nominal L1T rate of 100 kHz. Tracking algorithms are widely used in the HLT in the object reconstruction through particle-flow techniques as well as in the identification of b-jets and lepton isolation. Reconstructed tracks are also used to distinguish the primary vertex, which identifies the hard interaction process, from the pileup ones. This task is particularly important in the LHC environment given the large number of interactions per bunch crossing: on average 25 in 2012, and expected to be around 40 in Run II with a large contribution from out-of-time particles. In order to cope with tougher conditions the tracking and vertexing techniques used in 2012 have been largely improved in terms of timing and efficiency in order to keep the physics reach at the level of Run I conditions. We will present the performance of these newly developed algorithms, discussing their impact on the b-tagging performances as well as on the jet and missing transverse energy reconstruction.

## 1. Introduction

CMS [1] has a wide physics program for Run II (re-discovery of the Standard Model at 13 TeV, search of possible new physics, precision measurements of rare processes), therefore the main goal of the CMS trigger system is to keep the largest possible number of interesting events for analyses while keeping the event rate within the system limitation, namely 1 kHz.

One of the key ingredients is to make a wider use of the tracking and particle-flow based techniques. In addition, the identification of b-quark jets plays an important role in reducing the otherwise overwhelming processes involving jets from gluons (g), light flavour quarks (u, d,s) and from c-quark fragmentation.

In 2015, data taking operations are expected to restart at a centre-of-mass energy of 13 TeV with an instantaneous luminosity which should reach the peak value of  $1.4 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . In such conditions we are expecting an increase of event rate of about a factor of 4 with respect to the last period of data taking in 2012. Moreover, the expected average number of overlapping proton-proton interactions (pile up, PU) will be around 40. In these conditions the CMS tracker



is crossed by thousands of charged particles in each bunch crossing. In such a high occupancy environment designing tracking algorithms with high efficiency and a low fraction of fake tracks is very challenging. In addition the tracking code must run sufficiently fast so that it can be used at the HLT.

During the two years long period of maintenance and upgrade activities, many improvements have been developed in order to fulfill such requirements.

## 2. The CMS trigger system

The collision rate at the Large Hadron Collider (LHC) is heavily dominated by large cross section QCD processes, which are not of prime interest for the physics program of the CMS experiment. The processes relevant for new physics usually occur at a rate smaller than 10 Hz. Since it is not possible to record all the events and to select them later on, because of a limited bandwidth, it becomes mandatory to use a trigger system in order to select events according to physics-driven choices.

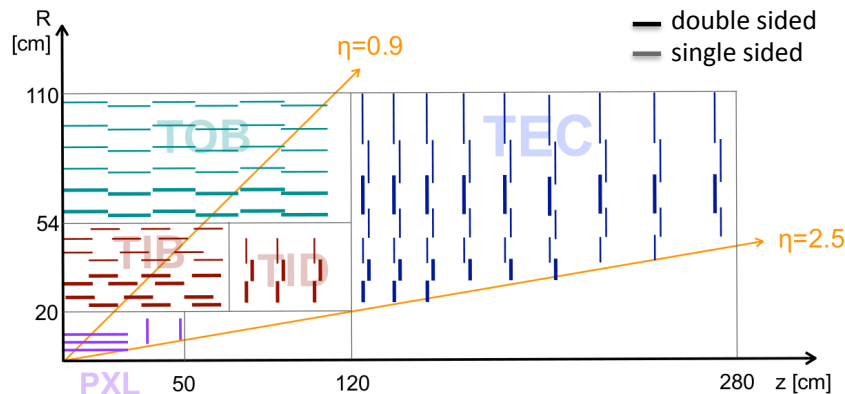
The CMS experiment features a two-level trigger architecture. The Level 1 Trigger (L1T) is an hardware level implemented on FPGA and custom-designed ASICs. The goal is to select events at maximum rate of about 100 kHz from the 40 MHz incoming rate delivered by the LHC. This upper limit is given by the CMS data acquisition electronics. L1T treats in parallel information coming from the electromagnetic and hadronic calorimeters as well as from the muons chambers. A global decision is taken based on the presence of energy deposits compatible with the presence of photons, electrons, muons, jets or hadronically decaying tau leptons.

The second level, called High Level Trigger (HLT), is a streamlined version of the CMS offline reconstruction software running on a cluster of commercial rack-mounted computers which consists of about 13,000 processors. The goal is to reduce the rate to about 1000 Hz for offline data storage.

## 3. The tracking system

The silicon tracking system, shown in Figure 1, is immersed in a magnetic field of 3.8T and is composed of a pixel silicon detector and a silicon micro-strip one. The pixel detector consist of three barrel layers at radii between 4.4 cm and 10.2 cm and two endcap disks at each side and sensors feature single pixel size of  $100 \times 150 \mu\text{m}^2$  for a total of 66M channels. The strip tracker covers the radial range between 20 cm and 110 cm around the LHC interaction point. The barrel region ( $|z| < 110\text{cm}$ ) is split into a Tracker Inner Barrel, made of four detector layers, and a Tracker Outer Barrel, made of six detector layers. The TIB is complemented by three Tracker Inner Disks per side. The forward and backward regions ( $120\text{ cm} < |z| < 280\text{ cm}$ ) are covered by nine Tracker End-Cap disks per side, thus extending the overall acceptance to cover the region  $|\eta| < 2.5$ . In some of the layers and in the innermost rings, special double-sided modules are able to provide accurate three-dimensional position measurement of the charged particle hits. The strip tracker is instrumented by about 15,000 modules with different strip pitches ranging from 80 to 180  $\mu\text{m}$ , for a total 9.6 million channels[1][3][4].

The basic performance of the tracking detector is a transverse momentum resolution  $\sigma(p_T)/p_T$  around 1-2% for muons of  $p_T$  around 100 GeV, an impact parameter resolution of 10-20  $\mu\text{m}$  for tracks with  $p_T = 10\text{-}20\text{ GeV}$  and the reconstruction of tracks belonging to a jet has an efficiency of about 85-90% and a few percent fake rate. Because the silicon strip unpacking is known to take a long time and to have a strong dependence on strip occupancy, regional and *on-demand* unpacking is performed at HLT, in which only modules requested during the pattern recognition are actually unpacked. The strip modules are grouped into geometrical regions, defined by a grid on the  $\eta$ - $\phi$  plane (where  $\eta$  is the pseudorapidity and  $\phi$  the azimuthal angle) with configurable dimensions (typically  $0.5 \times 0.5$ ), and raw data is considered only from regions-of-interest. More specially, any raw data packet with at least one channel connected to a region-of-interest is

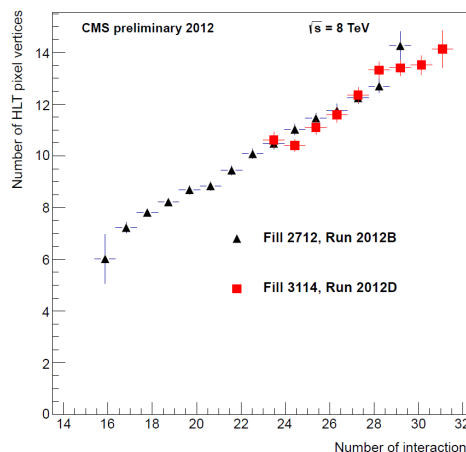


**Figure 1.** A simplified sketch of a quadrant of the R-z section of the CMS Tracker (bold lines represent double sided module assemblies).

fully unpacked. These regions-of-interest can be defined by physics objects identified in external sub-detectors (as muons, electrons, jets and taus). This is not used in the offline reconstruction as the track reconstruction searches the entire  $\eta$ - $\phi$  region and therefore needs all hits.

#### 4. Track reconstruction at HLT

The physics goals of CMS place strong requirements on the performance of the tracking. Searches for high mass resonances require tracks to have good momentum resolution for transverse momenta,  $p_T$ , of up to 1 TeV. Efficient reconstruction of very soft tracks with  $p_T < 1$  GeV is needed for studies of hadron production rates and to obtain optimum jet energy resolution with Particle Flow techniques [9]. It must be possible to resolve very close tracks, such as those from 3-prong tau decay. Excellent impact parameter resolution is needed for a precise measurement of the primary vertex position and for b-jet identification. The CMS tracker was designed in order to satisfy these requirements and the track finding algorithms are designed to fully exploit its capabilities and deliver the desired performance.



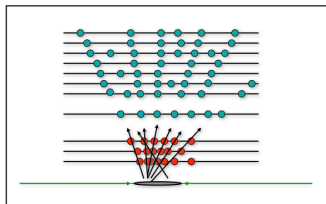
**Figure 2.** Number of reconstructed pixel vertices as function of the number of pile-up interactions.

The HLT uses track reconstruction software that is almost identical to that used for offline reconstruction [8], but it has to fulfill the CPU timing constraint: the event selection has to be done in about 150-200 ms. Tracks are reconstructed using hits from both pixel and strip detectors, but they can be reconstructed from hits found using only the pixel tracker as well. This is extremely fast, and is used with great effect in the reconstruction of the primary-vertex

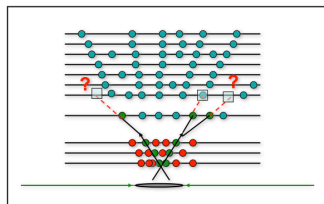
position. In 2012, for instance, the number of reconstructed vertices shows a linear dependence on the number of interactions without saturating (see Figure 2).

In CMS, tracks are reconstructed in four steps:

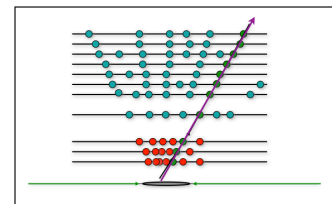
- the seed generation (*seeding*) provides initial track candidates (see Figure 3), defining the initial estimate of the trajectory parameters and their uncertainties.
- the pattern recognition (*building*), when track candidates are propagated using a Kalman filter technique [5] to find new compatible hits and the track parameters are updated, as shown in Figure 4.
- the final track fitting (*fitting*) is used to provide the best estimate of the parameters of each trajectory combining all the associated hits by means of a Kalman filter and smoother (see Figure 5).
- the track selection sets quality flags based on a set of cuts sensitive to fake tracks, on the track normalized  $\chi^2$ , and on its compatibility with interaction region.



**Figure 3.** Seeding.



**Figure 4.** Building.



**Figure 5.** Fitting.

Reconstruction efficiency relies on several iterations of the tracking procedure; each step, except for the first one, works on the not-yet-associated hits surviving the previous step, reducing the combinatorial complexity. In the early iterations tracks with relatively high  $p_T$ , produced near the interaction region, are reconstructed, discarding hits associated with those tracks, later iterations can search for lower  $p_T$  or highly displaced tracks. This procedure is referred to as *Iterative Tracking*.

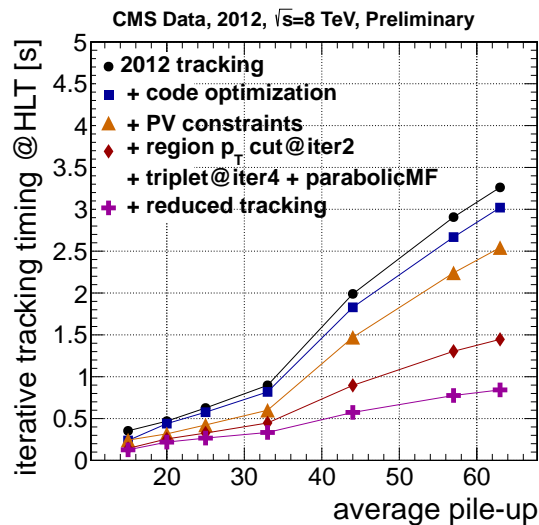
Because the tracking is a sophisticated and complex software and it is one of the most time consuming step (about 20% of the total CPU time), the HLT has to run it much faster. This is achieved by using a modified configuration of the track reconstruction, in particular by

- performing the track reconstruction only when necessary, and only at the end of the event selection process (after having applied other requirements based on the fast reconstruction of the physics object, as muon, electron, jet and tau);
- using a regional track reconstruction, where tracking is done only within regions-of-interest defined by the direction of the already available physics object;
- increasing the  $p_T$  requirement when forming the seeds (usually  $\sim 1$  GeV);
- selecting only the track phase-space in which tracks mostly come from the primary interaction region;
- stopping the track candidate building when specifics condition are met, for example, number of hits (typically 8), or the track parameters precision;
- limiting the maximum number of built candidates from a given seed (typically 2);
- limiting the number of iterations.

Moreover, because the silicon strip unpacking takes a long time and has a strong dependence on strip occupancy, regional and *on-demand* unpacking is performed at HLT, in which only modules requested during the pattern recognition are actually unpacked.

In 2015 the *Iterative Tracking* will consist of 4 iterations at HLT. The main differences between the 4 iterations lie in the configuration of the seed generation and final track selection steps. Iteration 0 reconstructs the most part of the tracks (around 80%) and is designed for prompt tracks, using the already reconstructed pixel tracks as seeds. Iteration 1 is configured to find low  $p_T$  prompt tracks and it is seeded by pixel triplets. Iteration 2 is used to recover prompt tracks which have only two pixel hits or slightly lower  $p_T$ . Iteration 4 is intended to find displaced tracks with respect to the beamspot. The last iteration is run only when highly displaced tracks are needed.

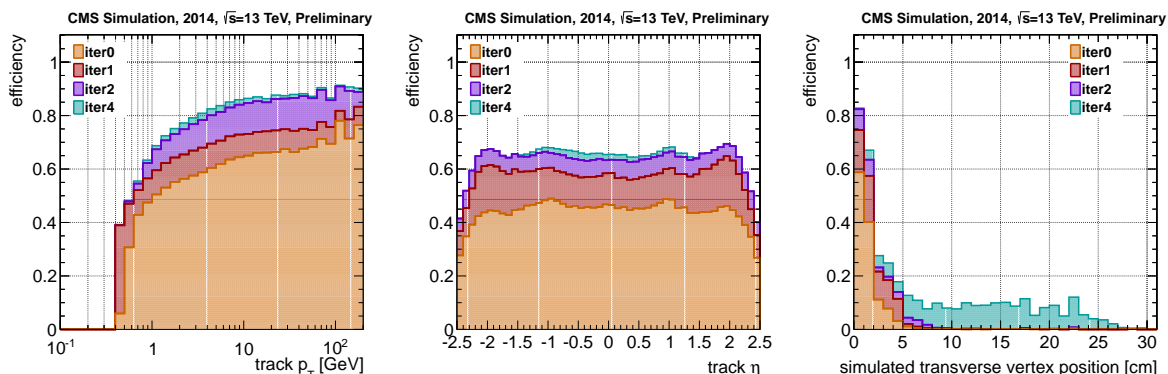
A factor 3.5 of improvement in the CPU time at  $\langle \text{PU} \rangle \sim 40$  has been obtained with respect to the 2012 tracking configuration by optimizing the *Iterative Tracking* at HLT, as shown in Figure 6.



**Figure 6.** Tracking timing per event vs average pile-up. The black curve refers to the tracking configuration used in 2012, while the other distributions refer to different tracking configurations, adding sequentially the changes foreseen for 2015. The purple distribution shows the final configuration designed for Run II operations.

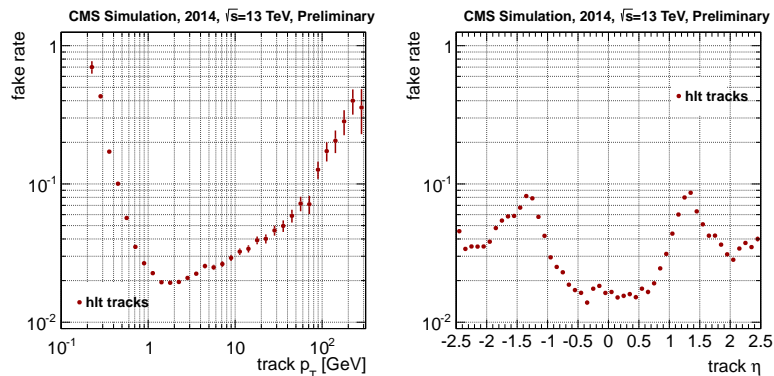
#### 4.1. Tracking performance

The performance of the *Iterative Tracking* algorithm at HLT has been evaluated on simulated  $t\bar{t}$  events at  $\sqrt{s} = 13$  TeV, with average pile-up 20 and bunch spacing 25 ns. The tracking efficiency is defined as the fraction of simulated charged particles that can be associated to a reconstructed track. It depends not only on the quality of the track finding algorithm, but also on the intrinsic properties of the tracker, such as its geometrical acceptance and material budget. The amount of material that a particle has to cross in the silicon tracker volume is far from being negligible (from  $0.4 X_0$  at  $\eta \sim 0$  to  $1.7 X_0$  at  $\eta = 1.5$ ). This cause a sizeable amount of photon conversions, bremsstrahlung and nuclear interactions in the tracker material. Figure 7 shows the track reconstruction efficiency as function of the main kinematics variables for each iteration, where the different phase-space of tracks reconstructed by each iteration can be clearly appreciated. The overall tracking efficiency at HLT, is around the 70%, while in offline reconstruction it is above the 90%.



**Figure 7.** Tracking efficiency as function of  $p_T$  (left),  $\eta$  (centre) and the transverse distance from the beam axis to the production point of each particle (right).

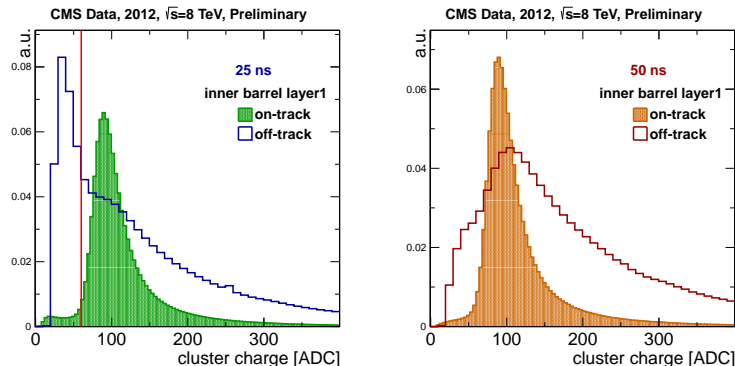
The fake rate is defined as the fraction of reconstructed tracks that are not associated with any simulated particle. This quantity represents the probability that a track produced by the reconstruction algorithm is either a combination of unrelated hits or a genuine trajectory that is badly reconstructed by including a large number of spurious hits. Figure 8 (right) shows the



**Figure 8.** Fake-rate as function of  $p_T$  (left) and  $\eta$  (right).

fake rate as function of the track  $\eta$ . As expected, the largest tracking fake rate comes from those regions of the tracker where the material budget is large. This effect is more significant for low energy hadrons due to their higher cross section for nuclear interactions, as shown in Figure 8 (left).

In addition, with 25 ns bunch crossing, out of time charged particles increase the occupancy of the detector; the effect is much larger for strips than for pixels. The increase in occupancy for the strip detector due to out of time particles induces a 2x increase both on timing and fake rate. Given that such particles cross the sensor at random time, the corresponding clusters are characterized by low collected charge. Figure 9 shows the typical cluster charge distribution collected by one layer of the strip detector per event. Cutting on the cluster charge, accounting for both for sensor thickness and the trajectory crossing angle, allows to reduce by  $\sim 50\%$  both the fake rate and the timing of the track reconstruction seeded by the strip hits.



**Figure 9.** Cluster charge collected by the strip detector inner barrel layer1 per event. Solid lines represent the distribution of clusters not associated to any tracks, while filled histograms to clusters which belong to a reconstructed track. Left: cluster distributions in 25ns bunch crossing events; right: events with 50ns bunch crossing. Red line represents the typical value of the cluster charge cut.

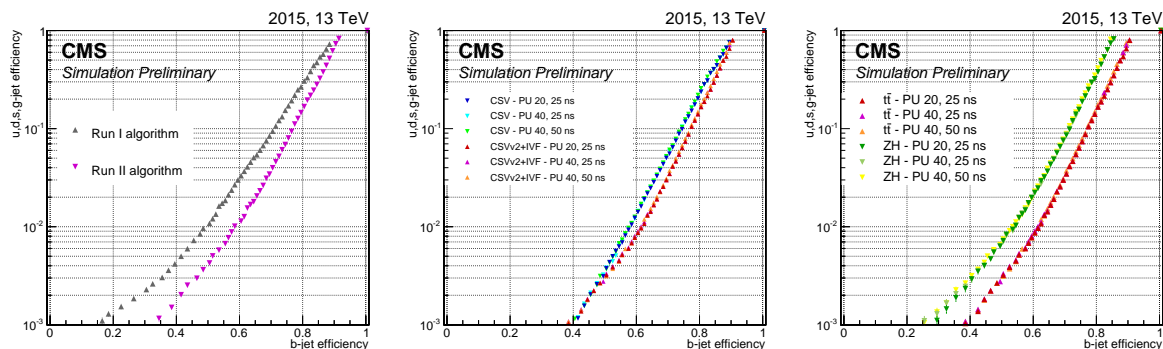
## 5. b-tagging at HLT

Many channels studied in the CMS physics program involve b-jets. Top quark, beyond the Standard Model and Higgs boson physics strongly rely on the capability to identify jets arising from the hadronization of b-quarks. A good performance in terms of b-jet identification efficiency and rejection of non-b-jets is essential to improve the purity of the selected events. Moreover, the use of b-tagging at trigger level is helpful –if not mandatory– to select efficiently the relevant events at trigger level while keeping the background sufficiently low to accommodate the online processing and bandwidth constraints. B-tagging at HLT is indeed one of the key ingredients to reduce the bandwidth when selecting hadronic events where the thresholds on the jet kinematics are not sufficient to reduce the background.

B-tagging algorithms exploit the properties of the B-hadrons such as their large decay life-



times ( $c\tau \sim 450 \mu\text{m}$ ) or the presence of leptons in the final state to separate them from light hadrons (made of u, d, s-quarks). Several algorithms have been developed and optimized, and they are based on variables such as the impact parameter of charged particle tracks, properties of reconstructed secondary vertices from heavy hadron decays, and the possible presence of a lepton in the jet, or combinations thereof. The algorithm which has been shown to guarantee the best performance in terms of b-jet identification efficiency and light-jets fake rate in Run I is the so-called Combined Secondary Vertex (CSV) algorithm [10]. It is a sophisticated algorithm which makes use of multivariate techniques to combine discriminating variables built from displaced track and secondary vertex information as well as jet kinematics. The offline b-jet identification efficiency is about 70% for a misidentification probability for light-jets of about 1% for jets with a transverse momentum  $p_T$  between 80 and 120 GeV. During the long shutdown the CSV algorithm has been significantly improved (CSVv2) by updating the multivariate algorithm from a simple Likelihood Ratio method to a more sophisticated Neural Network one, by improving the track selection and adding new variables, and by using a new algorithm for the reconstruction of the secondary vertices, the so called Inclusive Vertex Finder (IVF). These updates turn into a net improvement in performance: about 10% increase in the b-jet identification efficiency at a 1% of light-jets misidentification rate.



**Figure 10.** Efficiency of the b-tagging algorithm used at HLT for u, d, s, g-jets vs b-jets at  $\sqrt{s}=13$  TeV for different pile-up and bunch spacing scenarios. Left: comparison between Run I and Run II configuration on a  $t\bar{t}$  sample. Centre: comparison between the Run I algorithm (CSV) and its improved version for Run II (CSVv2+IVF) on a  $t\bar{t}$  sample. Right: comparison between  $t\bar{t}$  (large jet multiplicity) and  $ZH(\rightarrow b\bar{b})$  (small jet multiplicity) samples.

The b-jet identification at HLT relies on the same version of the algorithm used in the offline reconstruction. The b-tagging algorithm at HLT has been revisited and optimized for the Run II, with the main goal of reducing the execution time. This was achieved thanks to a more extensive use of the *Iterative Tracking* reconstruction in the regions-of-interest, which helps in reducing the combinatorics. The second goal is to improve the performance of the online version of the b-tagging algorithm. This was achieved thanks to the *Iterative Tracking* reconstruction, to the new implementation of the b-tagging algorithm (CSVv2+IVF) and to a better resolution on the position of the primary vertex. By using tracks from the *Iterative Tracking*, and performing the vertex fitting with an Adaptive Vertex Fitter [6] where the track clustering is made with a Deterministic Annealing (DA) algorithm [7], indeed, the resolution on the position of the primary vertex along the z-axis is about 20-30  $\mu\text{m}$ .

Figure 10 (left) shows the achieved improvements with respect to the 2012 configuration in terms of the b-jet efficiency as function of the light-jets one. An improvement of about 10% is obtained for a fake-rate of 1%. Figure 10 (centre) shows the comparison in terms of performance between the two versions of the b-tagging algorithm, CSV used in Run I and CSVv2+IVF developed for Run II operations.

In addition, in order to speed up the identification of the main interaction region, where the

primary vertex lies, and from where the *Iterative Tracking* starts to look for seeds, a dedicated algorithm has been developed. The goal is to determine the position of the vertex along the beam axis (z-axis), starting from the jet direction, and using only pixel clusters information. The so called Fast Primary Vertex (FastPV) reconstruction starts by selecting pixel clusters with the same  $\phi$  coordinate as the jet, and with size along the y-axis compatible with the pseudo-rapidity  $\eta$  of the jet (the pixel cluster size increases for jet at high  $\eta$ ) and small size along the x-axis (small pixel size selects high  $p_T$  tracks). Then, the selected pixel clusters are projected along the z-axis, and the resulting peak in the z direction, where most of the clusters ended up, is used as an estimate of the primary vertex position along the beam axis.

The tracks reconstruction used by the b-tagging algorithm relies on this FastPV, therefore the performance of the b-tagging shows a dependence on the jet multiplicity in the events. This is shown in Figure 10 (right), where the b-tagging performance on a sample of  $t\bar{t}$  and a  $ZH(\rightarrow b\bar{b})$  one are compared. The measured performance in  $t\bar{t}$  events shows a b-jet efficiency of about 60% for a misidentification probability of about 1%, while, for the same misidentification probability, the  $ZH(\rightarrow b\bar{b})$  sample has a slightly lower efficiency.

## 6. Conclusion

CMS plans to extend the usage of particle-flow technique at HLT in Run II and to use the tracking also in lepton isolation and b-tagging in order to improve the signal efficiency and background rejection. For such algorithms, high efficiency and low track fake rate are the key ingredients, therefore the *Iterative Tracking* procedure applied in the region-of-interest is the best choice.

At HLT the track reconstruction is done by using the same algorithm as in the offline reconstruction, based on Kalman filter techniques. The need of having the highest possible performance, in terms of high track efficiency and low fake-rate, while keeping the CPU timing within the constraint, led to the development of an *ad hoc* tracking configuration at HLT. This is able to reconstruct tracks over the full rapidity range of the tracker. For promptly produced charged particles the average tracking efficiency is typically 70%. This performance is not as high as the offline version, but it has been shown to guarantee a good performance in terms of the physics objects. The application of the *Iterative Tracking* at HLT allows to decrease the event rate while keeping a high efficiency on selecting events for the physics analysis. During the two years of shutdown many developments have been done in order to speed up the code while guaranteeing the highest efficiency as possible. A reduction in the tracking timing of about a factor of 4 with respect to the Run I setting has been achieved with similar trigger performance at  $\text{PU} \sim \langle 40 \rangle$ .

The b-jet identification algorithm has been successfully used at HLT during Run I for many analyses, in particular those involving  $b\bar{b}$  resonances such as the Higgs boson decaying to  $b\bar{b}$ . Improvements with respect to the processing time and the performance have been developed leading to a promising use during Run II. The improvements are mainly due to a multi-step approach based on an intensive usage of regional reconstruction techniques.

- [1] CMS Collaboration, *JINST* **3**, S08004 (2008).
- [2] CMS Collaboration, CMS The TRIDAS Project, TDR CERN LHC 02-26, CMS TDR 6 (2002).
- [3] CMS Collaboration, The Tracker System Project TDR CERN-LHCC 98-6 (1998).
- [4] CMS Collaboration, Addendum to the CMS Tracker TDR CERN-LHCC 2000-16 (2000).
- [5] R. Fruhwirth, Nucl.Instrum.Meth. A262 (1987) 444-450.
- [6] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, CMS Note 2007-008, 2007.
- [7] K. Rose, E. Gurewitz, G. Fox, Proceedings of the IEEE 86 (1998) doi:10.1109/5.726788.
- [8] CMS Collaboration, CMS-TRK-11-01, Submitted to JINST (2014).
- [9] CMS Collaboration, CMS-PAS PFT-09-001(2009).
- [10] CMS Collaboration, CMS-PAS-BTV-13-001 (2013).