Performance of track and vertex reconstruction and b-tagging studies with CMS in pp collisions at $\sqrt{s} =$ 7 TeV

Alexander Schmidt on behalf of the CMS Collaboration[*]
Universität Zürich, CH-8057 Zürich, Switzerland
E-mail: Alexander.Schmidt@cern.ch

First 7 TeV proton-proton collisions produced by the Large Hadron Collider at CERN have been recorded by the CMS experiment in 2010. The ability to accurately and efficiently reconstruct the trajectories of charged particles produced in these collisions is a critical component for most measurements at the LHC. Several methods for determining the efficiency and measuring the resolution of track parameters using the first data collected by the detector are presented. The trajectories of charged particles are the key elements for the reconstruction of the primary interaction vertex and for the identification of jets containing the products of a B-hadron weak decay. The distributions of the key observables, such as track impact parameters and vertex properties, are compared with the prediction of Monte Carlo simulation. Finally the performance of vertex reconstruction and B identification algorithms are derived with data-driven methods.

---

[*]Speaker.

1. Introduction

The Compact Muon Solenoid (CMS) [1] is one of two general purpose detectors at the Large Hadron Collider (LHC) at CERN. Its main purpose is to search for new physics at highest energies. CMS consists of a tracking detector (described in Section 2), electromagnetic and hadronic calorimeters and a muon system. Among the most fundamental objects for the reconstruction of physics processes in proton-proton collisions are the tracks of charged particles which are measured with the tracker. From these tracks higher-level objects such as vertices and b-jets can be constructed. This paper summarizes the performance of the tracker and gives an overview of track reconstruction, vertex reconstruction and b-jet tagging.

2. The CMS Silicon Tracker

The CMS tracking detector [1] is an all-silicon device. The innermost part consists of silicon pixel layers (three barrel layers and two endcap disks per side), followed by silicon strip layers (ten barrel layers and twelve endcap disks per side). The pixel barrel covers a radial region between 4 cm and 15 cm and a longitudinal region between $\pm$ 50 cm. The strip tracker covers the region from 25 cm to 110 cm in radius and within 280 cm along the beam axis. The tracker acceptance extends to a pseudo-rapidity of $|\eta| < 2.5$.

With first LHC proton collisions it was possible to measure the material budget and to compare it with the predictions of the Monte Carlo simulation. The measurement of the material budget has been performed using different independent approaches, for instance by reconstructing the amount and position of photon conversions or nuclear interactions. The agreement between data and simulation has been found to be within 10%. An illustration of the photon conversion method is shown in Figure 1 (left) where the distribution of reconstructed photon conversion vertices is displayed for the pixel barrel detector. This figure shows a visible displacement of the pixel barrel detector with respect to the beam pipe.

3. Track reconstruction and tracking performance

The reconstruction of trajectories of charged particles is accomplished in an iterative procedure using the "Combinatorial Track Finder" [3]. Each iteration starts with the creation of seeds from which the trajectory is propagated through the full tracker. The found tracks are cleaned to remove fake tracks. The hits attached to the found tracks are removed from the hit collection and the procedure is repeated with the remaining hits. There are six iterations in total. The main difference between the iterations is the seeding step. In the first two iterations pixel pairs and triplets are used as seeds to find prompt tracks with a momentum threshold of $p_t > 0.9$ GeV. The third iteration removes the momentum cut, while the fourth iteration includes hits from the strip tracker to reconstruct displaced tracks. The final two iterations use seeds of strip pairs to reconstruct tracks lacking pixel hits, for instance to identify photon conversions.

The tracking efficiency is measured with various methods [4]. In the first method, hits from simulated tracks are embedded in the data. This has the advantage that it can be done with all types of tracks and momentum regions. There are also purely data driven methods, for example
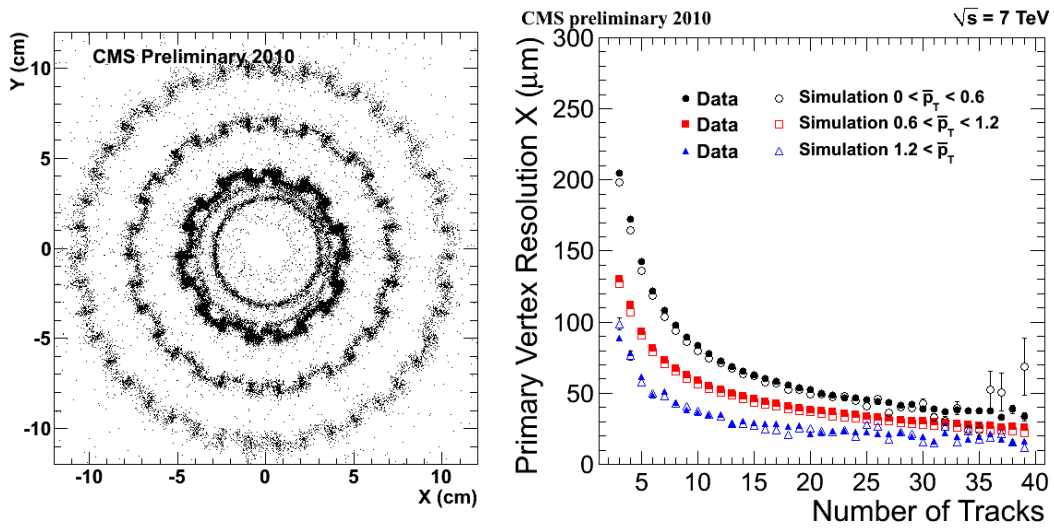
Figure 1: Left: radiography of the CMS pixel barrel detector using reconstructed photon conversion vertices. The innermost circle is the beam pipe. Right: primary vertex position resolution depending on the number of tracks for three different regions of average track transverse momentum (p_t).

the tag-and-probe technique which uses di-muon resonances (e.g. J/ψ mesons). In this method, a high-quality muon is used as the "tag" while the "probe" muon is reconstructed using information from the muon system only. Passing or failing probes are defined as those that are matched or not matched to tracks in the silicon tracker. The yield of passing and failing probes is then measured using an unbinned maximum likelihood fit to the di-muon invariant mass distribution in the J/ψ region. The obtained efficiency is corrected for the matching efficiency and fake (random) matches. The results show that the tracking efficiency is between 97.6% in the forward region and 100% in the barrel region. The prediction from the Monte Carlo simulation agrees with data at the one percent level.

4. Primary vertex

The primary proton-proton interaction vertices are reconstructed from prompt tracks which are selected using the transverse impact parameter with respect to the beamspot. These prompt tracks are then clustered along the *z*-axis by requiring a separation of at least 1 cm to the next cluster. Finally the clusters are fit using an adaptive vertex fitter [5] in which outliers are downweighted.

A data-driven method is applied to measure the primary vertex reconstruction efficiency and resolution at the same time [6]. In this method the tracks at the primary vertex are randomly split into two sets. Both sets are independently fit with the procedure described above. The resolution is then extracted by fitting a single gaussian to the distribution of the difference of the two vertex positions. The primary vertex resolution strongly depends on the number of tracks used in the vertex fit and on the transverse momentum. Figure 1 (right) shows the primary vertex position resolution depending on the number of tracks for three different regions of the average transverse momentum. The comparison with the Monte Carlo simulation shows very good agreement.
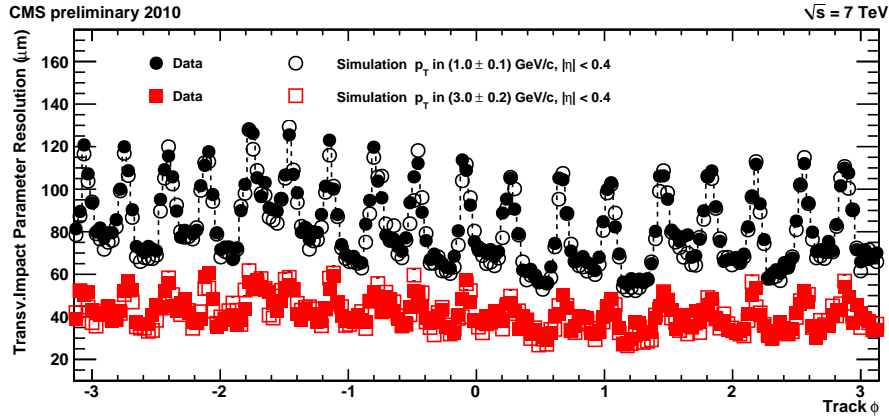
Figure 2: Resolution of the transverse impact parameter depending on the azimuthal angle ϕ for two different track p_t ranges. The "oscillating" structure is due to the cooling pipes of the inner layer of the pixel detector.

In order to measure the primary vertex efficiency, one set of split tracks is considered to be a "tag" vertex while the other one is the "probe". The efficiency is then calculated as the fraction of reconstructed probes matching the tag vertex. The vertices are considered to be matched if their distance in z is smaller than 5σ. The result is a measured efficiency of almost 100% if there are more than two tracks with transverse momenta greater than 0.5 GeV in the vertex.

5. Impact parameter

The impact parameter (IP) is defined as the distance between the primary vertex and the track at their point of closest approach [7]. Prompt tracks from the primary vertex have small IP values while tracks from decays of B hadrons have rather large IP values because of the B hadron lifetime which leads to flight distances of several millimeters. The IP is therefore an important quantity to identify B hadron decays in jets ("b-jet tagging").

The IP resolution of prompt tracks is measured with a data driven method [6]. In this method the tracks from a primary vertex are used and the measured track is removed from this list of tracks. The primary vertex fit is then redone without the track to be measured. The IP of this track is then a gaussian distribution, centered around zero. The width of this distribution is a convolution of the IP resolution and the PV resolution. The IP resolution is then extracted by fitting convoluted gaussians. This resolution depends on track momentum and pseudo-rapidity and reaches 20 μm in the barrel for tracks with $p_t > 12$ GeV. Figure 2 shows the transverse IP resolution depending on the azimuthal angle ϕ for two different p_t ranges. The "oscillating" structure in this figure is due to the cooling pipes of the inner layer of the pixel detector as visible in Figure 1 (left). Another visible structure in Figure 2 is a global modulation with ϕ. The region around $\phi = 1$ shows better resolution than the other regions which is due to the displacement of the pixel detector which is also visible in Figure 1. A third "oscillation" in Figure 2 is due to the sensors being closer or further from the beam line due to the detector geometry with alternating module positions.
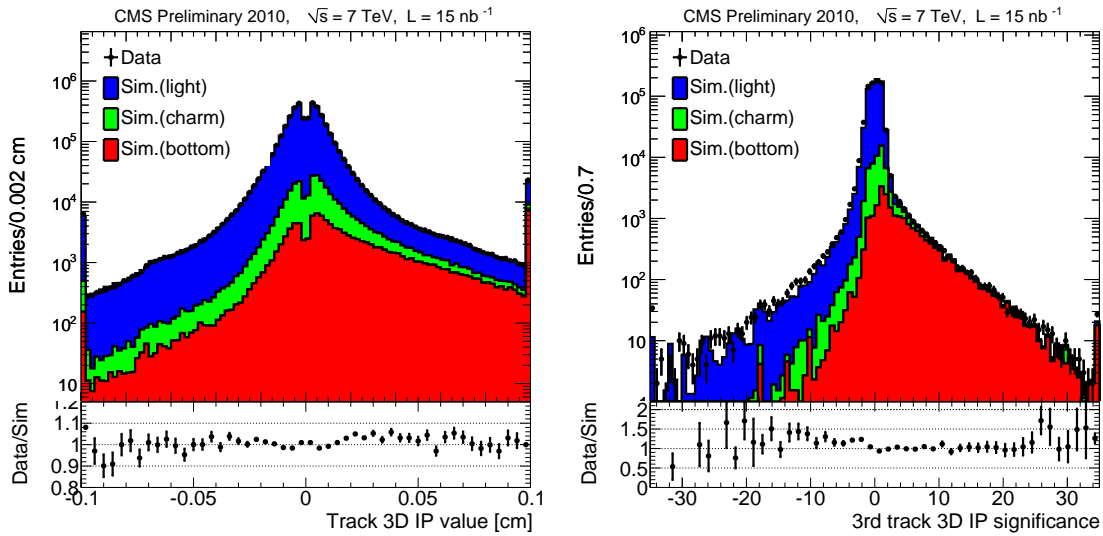
Figure 3: Left: impact parameter value of all selected tracks in a jet. Right: impact parameter significance of the third track in a jet (ordered by IP significance). The Monte Carlo simulation of light, charm and b-jets is shown in blue, green and red, while the data are represented as black markers.[8]

The distribution of the IP for all selected tracks in a jet is shown in Figure 3 (left). The used quantity in b-tagging algorithms is the IP significance which is defined as the IP value divided by its error. This definition reduces the impact of mismeasured tracks having large IP values. In the "track counting" b-tagging algorithms [9], the tracks are ordered by their significance and the second or third track IP significance is used as discriminator. Figure 3 shows the IP significance distribution of the third track in a jet. It is visible that the positive tail of this distribution contains almost only b-jets.

6. Secondary vertex

Secondary vertices in jets are reconstructed using the adaptive vertex fitter [5]. The presence of a secondary vertex with a certain number of tracks is already a strong hint for the presence of the decay of a long-lived particle such as a B hadron. This can be seen in Figure 4 (left) which shows the average number of reconstructed tracks at the secondary vertex depending on the jet momentum.

The secondary vertex allows to calculate several important quantities that distinguish b-jets from other jets. An important quantity is the flight distance, which is the separation between the primary and the secondary vertex. It is used in the secondary vertex b-tagging algorithms to calculate a discriminator [9]. The secondary vertex discriminator is shown in Figure 5 (left). Another crucial quantity is the invariant mass of the secondary vertex which is calculated assuming the pion mass for each track. Figure 4 (right) shows the distribution of the secondary vertex mass. It is clearly visible that this distribution allows to distinguish between light flavours, charm and beauty. It is therefore commonly used to determine the purity of a b-tagged sample, for example in
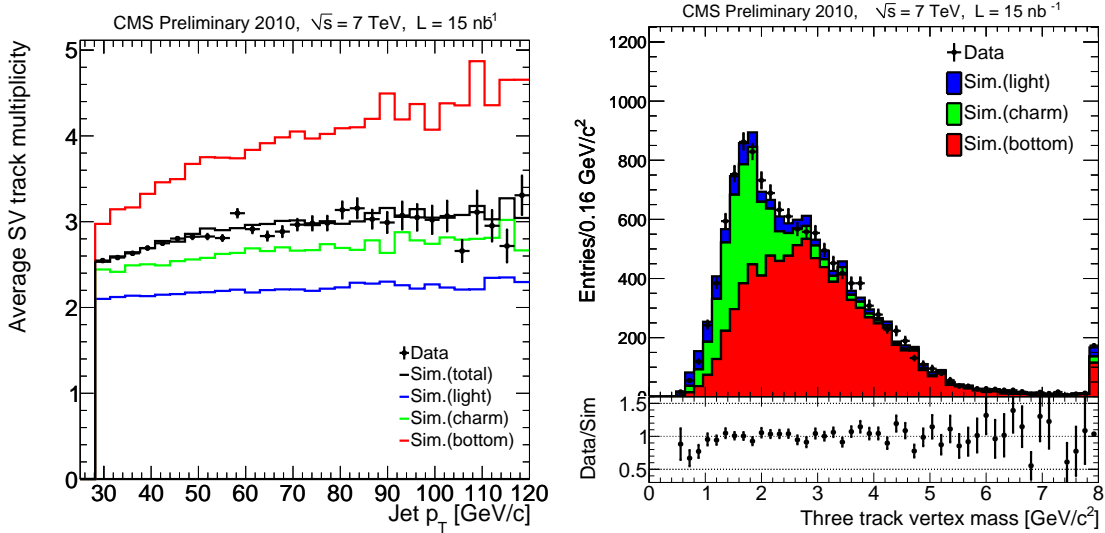
Figure 4: Left: average number of tracks at the secondary vertex depending on the jet transverse momentum. Right: invariant mass of secondary vertices with at least three tracks. The Monte Carlo simulation of light, charm and b-jets is shown in blue, green and red, while the data are represented as black markers.

the measurement of the inclusive b-production cross-section [10] which was the first CMS precision measurement using b-tagging.

7. b-tagging performance

The measurement of the b-tagging efficiency has been done using muons in jets [8]. About 10% of the B decays contain a muon in the final state (20% if muons from charm decays in the $b \rightarrow c$ cascade are taken into account). The relative momentum of the muon with respect to the jet (p_t^{rel}) is a variable that allows to distinguish between true B decays and background. This is due to the high mass of B mesons of around 5 GeV which causes a large muon momentum relative to the B flight direction. Templates for the p_t^{rel} spectra of light, charm and b-jets are obtained from Monte Carlo simulation. Light flavour templates can also be obtained directly from data, using "any" random non-muon track. The difference between the templates from simulation and data are small and are included as systematic errors.

The b-tagging efficiency is then obtained by fitting for the b-fraction in tagged and untagged samples using p_t^{rel} template fits. The detailed results for all tagging algorithms and working points can be found in [8]. The main result is the excellent agreement with the Monte Carlo simulation at the 10% level. However, the systematic error of this result is still around 20%. This uncertainty will be improved with more data and advanced methods.

The mistag efficiency is determined using the method of negative tags. In case of IP based b-tagging algorithms the discriminator gets a negative sign if the decay seems to have taken place on the other side of the primary vertex with respect to the jet direction. Mathematically, this is done by evaluating the scalar product of the IP segment (the vector from the primary vertex to the point of closest approach) with the jet direction. In case of secondary vertex algorithms this is
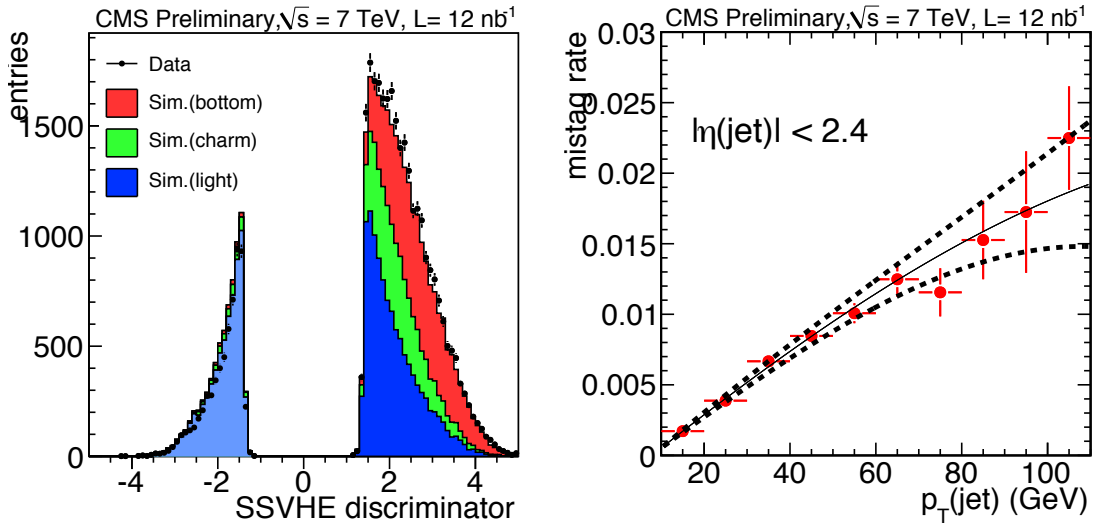
Figure 5: Left: discriminator of the secondary vertex b-tagging algorithm for negative and positive tags. Right: light flavour mistag rate as measured by the negative tag method depending on transverse jet momentum p_t for the secondary vertex b-tagging algorithm.

done in a similar way by deciding if the secondary vertex is in the same direction as the jet. Figure 5 (left) shows the distribution of the discriminator for negative and positive tags in case of the secondary vertex b-tagging algorithm. In the ideal case the discriminator is symmetric around zero for light flavour jets because any displacement is due to mismeasurements and has therefore equal probability to obtain a positive or negative sign. In addition, real b-jets should only have positive signs, because of the presence of a real displacement towards the jet direction. The amount of positive tags (mistags) of light flavour jets should therefore be identical to the total amount of negative tags. However, in reality, the distribution for light flavours is not perfectly symmetric because of long-lived particles such as the K_s. B-jets also have a small fraction of negative tags (especially for the IP b-tagging algorithms) because of resolution effects. The negative tag rate has therefore to be corrected for these effects using a scale factor R_{light} which is determined from Monte Carlo simulation. The results in terms of mistag efficiency are displayed in Figure 5 (right) for a secondary vertex b-tagging algorithm. It is visible that the mistag rate strongly depends on the transverse jet momentum p_t. The comparison of the measured results with the prediction from the Monte Carlo simulation shows an agreement between 70% for low-p_t jets (15 GeV) up to 90% for high-p_t jets (100 GeV). The detailed results for all algorithms and working points are given in [8].

8. Conclusions

CMS has been extremely successful in commissioning the tracking, vertex reconstruction and b-tagging tools with first collisions in 2010. The agreement with the Monte Carlo simulations was beyond expectations. The b-tagging tools have been applied right from the start in precision measurements such as the measurement of the inclusive b-production cross-section and even the

measurement of the top quark production cross-section. In the meantime the b-jet tagging is an indispensable tool in advanced searches for the Higgs boson and searches for physics beyond the standard model such as Supersymmetry. The work to improve the data-driven performance measurements and to reduce the systematic errors is ongoing.

References

[1] The CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 0803 (2008) S08004.

[2] The CMS Collaboration, "Studies of Tracker Material", CMS Physics Analysis Summary TRK-10-003 (2010).

[3] W. Adam et al., "Track Reconstruction in the CMS Tracker", CMS Note 2006-041 (2006).

[4] The CMS Collaboration, "Measurement of Tracking Efficiency", CMS Physics Analysis Summary TRK-10-002 (2010).

[5] R. Fruewirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", CMS Note 2007/008 (2007).

[6] The CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary TRK-10-005 (2010).

[7] A. Rizzi, F. Palla, G. Segneri, "Track impact parameter based b-tagging with CMS", CMS Note 2006/019 (2006).

[8] The CMS Collaboration, "Commissioning of b-jet identification with pp collisions at $\sqrt{s}$=7 TeV", CMS Physics Analysis Summary BTV-10-001 (2010).

[9] The CMS Collaboration, "Algorithms for b Jet Identification in CMS", CMS Physics Analysis Summary BTV-09-001 (2009).

[10] The CMS Collaboration, "Inclusive b-jet production in pp collisions at $\sqrt{s}$=7 TeV", CMS Physics Analysis Summary BPH-10-009 (2010).