Measurement of Tracking Efficiency

The CMS Collaboration

Abstract

This note describes several methods for measuring the efficiency to reconstruct charged particles in the inner tracker of the CMS detector at the CERN Large Hadron Collider. The first method obtains tracking efficiency for muons and pions by embedding simulated tracks in data. Two additional methods are used to measure the efficiency of tracking muons, isolated and non-isolated, while a fourth method exploits the ratio of four-body and two-body $D^0$ decays to measure tracking efficiency for pions. Results generally indicate that the efficiency is in agreement with expectations from detector simulation within percent uncertainty.
1 Introduction

The Compact Muon Solenoid (CMS) [1] is one of two general-purpose detectors operating at the Large Hadron Collider (LHC) facility at CERN. One of the central features of the CMS detector is a 6-m diameter solenoidal magnet operating at 3.8 T, which enables the measurement of charged particle momenta over more than four orders of magnitude, from less than 100 MeV/c to more than 1 TeV/c, by reconstructing their trajectories as they traverse the CMS inner tracking system. The tracker is composed of a pixel detector with three barrel layers at radii of 4.4, 7.3, and 10.2 cm, and a silicon strip tracker with 10 barrel detection layers covering radii between 25 and 110 cm. Each system is completed by endcaps on each side of the barrel, extending the acceptance up to a pseudorapidity of |η| < 2.5.

The precise and efficient determination of charged-particle momenta is a critical component of the physics program of CMS, as it impacts the ability to reconstruct leptons, charged hadrons, jets, and photon conversions, which are the basic physics objects needed to understand pp collisions at the LHC. Reconstruction of tracks in the inner tracker is seeded by the average beamspot position and hits in the pixel detector [2]. Pixel seeds are then propagated outward, adding compatible hits in the strip tracker and updating the trajectory until either the detector boundary is reached, or no additional compatible hits can be found. In the final stage, the collection of hits is fit to obtain the best estimate of the track parameters.

As the instantaneous luminosity achieved by the LHC continues to increase, CMS will experience different occupancy environments, with multiple interactions per beam crossing and dense track multiplicity in jets. No single method of measuring track reconstruction efficiency can cover all possible environments, kinematic ranges, and systematic effects, so it is important to explore multiple overlapping methods to account for differing event occupancies, particle momenta, and sources of systematic uncertainty. In this note, we present four methods for determining the efficiency of full track reconstruction (pixel+strips) in CMS. A track-embedding method is used to measure absolute tracking efficiency for muons and pions for transverse momenta above 500 MeV/c. Two methods are used to measure tracking efficiency specifically for muons: one using J/ψ decays covering momenta\(^1\) from 1.5 to 10 GeV/c, and a second using non-isolated muons in heavy-quark jets for muon momenta above 5 GeV/c. The fourth method measures the ratio of tracking efficiency in data and simulation for low transverse-momentum pions between 300 MeV/c and 1.5 GeV/c using two-body and four-body decays of neutral charm mesons. The methods are complementary, and are subject to different systematic uncertainties, which allows for an internal consistency check among the different results. By considering the comparison of data and simulated samples across the four methods, we derive systematic uncertainties for the tracking efficiency of muons and pions that can be applied to a wide range of measurements.

2 Track Embedding

Absolute tracking efficiency can be studied in data by embedding simulated tracks in data events. This technique makes use of a hit-embedding procedure in which the probability of correctly reconstructing a simulated track is measured both as an isolated track in a simulated event, and after merging the simulated detector response with the digitized tracker information in a pp collision. The advantages of this method include the ability to separate precisely the effects of acceptance and reconstruction efficiency, the unlimited momentum range and occupancy environments that can be sampled, and the relatively small data samples required to

---

\(^1\)With more data, this method will be extended to higher momenta using leptonic decays of Z bosons.
achieve high statistical precision.

2.1 Method

The general track-embedding method involves simulating an isolated charged particle using
the CMS simulation software based on GEANT4[3], reconstructing the track using the hits
generated in the silicon tracker, and then embedding these hits in a data event to determine
if a track matching the original is successfully reconstructed. For the purpose of this study,
we separate the total efficiency to reconstruct a charged particle in CMS into acceptance and
reconstruction efficiency. The acceptance is defined as the probability that a charged particle
produces a sufficient number of hits in the tracker to be reconstructed by the track-finding al-
gorithm, while the efficiency is defined as the probability that these hits are used to reconstruct
a track with parameters representative of those of the original particle. Separated in this way,
the acceptance includes effects such as detector geometry, material budget, and silicon sensor
performance, while the efficiency is sensitive to the track-finding algorithm and hit occupancy,
which depends on the multiplicity of low-momentum particles that are difficult to simulate
accurately.

In practice, the efficiency is defined as the probability of reconstructing an embedded simulated
track in a data event, given that it could be reconstructed as an isolated track in a simulated
event. The total probability of successfully reconstructing a track can then be expressed as

\[ P_{\text{track}} = A \cdot e = \frac{N_{\text{reco,iso}}}{N_{\text{gen}}} \cdot \frac{N_{\text{reco,embed}}}{N_{\text{reco,iso}}}, \]  

(1)

where \( N_{\text{gen}} \) is the number of simulated tracks, \( N_{\text{reco,iso}} \) is the number of correctly reconstructed
simulated tracks, and \( N_{\text{reco,embed}} \) is the number of correctly reconstructed tracks in data that
were also reconstructed in the simulation. We consider two independent metrics for deciding
whether a track should be classified as being correctly reconstructed or not. The primary
method defines a successfully reconstructed track as one where 75% of the hits used to recon-
struct the track were also associated with the simulated charged particle. As a cross-check, we
use a simple cone-based association, requiring that the direction of the reconstructed track
momentum at the production point lie within a cone of radius \( \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} = 0.01 \) around
the simulated track direction.

2.2 Results

Track reconstruction efficiencies were determined by embedding simulated isolated muons
and pions in a subset of 7 TeV collision data recorded with a minimum-bias trigger. Charged
particles were generated uniformly in azimuth within the pseudorapidity range \(|\eta| < 3.0\), and
at discrete values of \( p_T \). Figure 1 shows the results for acceptance and efficiency as a function of
the generated \( p_T \) of the embedded track, both for muons and pions satisfying \(|\eta| < 2.4\), which
corresponds to the detector boundary. As a cross-check, we also show on the same plots the
acceptance and efficiency for tracks embedded in simulated minimum-bias events. For both
muons and pions, the efficiency is approximately 0.5% lower when embedding tracks in data
compared to simulation.

3 Muon Tag-and-Probe

A standard method for measuring muon trigger and reconstruction efficiencies in data is the
so-called tag-and-probe technique, in which a dimuon resonance (such as \( J/\psi \)) is reconstructed
3.1 Method

Reconstruction of muons in CMS is described in detail elsewhere [4]. Standalone muons are reconstructed by fitting hits in the muon detector system only, while global muons combine a standalone muon with a silicon track in a global fit using hits from the muon system and the inner tracker. Probe muons are defined simply as all standalone muons. Tag muons are defined as global muons that pass the muon high-level trigger criteria [4] with a threshold of 3 GeV/c, and that have a minimum transverse momentum of 2.6 GeV/c as determined from the silicon track fit.

To determine the tracking efficiency for probe muons, we match standalone muons to candidate silicon tracks defined with the high purity selection described in [2]. Probe muons are defined...
as passing or failing depending on whether or not they are matched to tracks passing the above quality requirements. The criteria used to match a silicon track to a probe muon are $|\Delta \eta| < 0.2$ and $\Delta R < 0.5$, which results in a track-muon matching efficiency of nearly 100% and a fake matching rate of approximately 10%.

The yield in data of passing and failing probes is then measured using an unbinned maximum-likelihood fit to the dimuon invariant mass in the region of the $J/\psi$ meson for tag muons paired with passing and failing probe muons, respectively. For signal we use a single Gaussian function, while for the background we use a third-order Chebyshev polynomial. From the fitted yields for passing and failing probes, we determine the combined efficiency $\epsilon$ of reconstructing a track and matching it to the standalone muon. When calculating the dimuon invariant mass we use the track information for the tag muons, while for the probe muons we use only the standalone muon information.

The efficiency determined in this way can be expressed in terms of the true tracking efficiency $\epsilon_T$, matching efficiency $\epsilon_M$, and the probability of fake (random) matches $\epsilon_F$ as

$$\epsilon = \epsilon_T \epsilon_M + (1 - \epsilon_T \epsilon_M) \epsilon_F.$$  \hspace{1cm} (2)

The probability of fake matches can be measured directly in data using the following technique. Before applying the track-matching selection to the probe muon, we remove all silicon tracks that, when combined with the tag muon, give an invariant mass near the $J/\psi$ mass. The rate of probe muons passing the matching criteria after removing such tracks is then a measure of the probability of getting a fake match between probe and track. We then invert Eq. 2 to obtain the combined probability of reconstructing a track and matching it to a probe muon,

$$\epsilon_T \epsilon_M = \frac{\epsilon - \epsilon_F}{1 - \epsilon_F}.$$  \hspace{1cm} (3)

Given that the track-muon matching criteria have been defined to be fully efficient, we interpret the resulting probability as the muon tracking efficiency.

### 3.2 Results

The method has been performed using collision data corresponding to an integrated luminosity of about 125 nb$^{-1}$, and on a simulated Monte Carlo sample corresponding to 1 pb$^{-1}$. For the kinematic range of muons used in this analysis ($p_T > 1.5\text{ GeV}/c$), the tracking efficiency is expected to be independent of muon momentum. We therefore measure the tracking efficiency as a function of pseudorapidity only. Figure 2 shows an example for muons with $|\eta| < 1.1$ of the dimuon invariant mass for tag-probe combinations where the probe muons pass or fail the track-matching criteria.

Table 1 summarizes the corrected tracking efficiency measured in data and simulated samples, and the ratio, in four independent $\eta$ ranges. We find efficiencies of about 98% or higher for all detector regions. The contribution to the total uncertainty due to fake matching is negligible, indicating that the method is not highly sensitive to the matching criteria. Figure 3 shows the corrected tracking efficiency as a function of $\eta$ for data compared to the simulated sample. From these results, we determine a systematic uncertainty on the tracking efficiency for muons in Monte Carlo simulation of 1 – 2% depending on muon pseudorapidity.

With the recent increase in instantaneous luminosity achieved by the LHC, the number of primary vertices per proton-proton interaction has increased significantly in the later part of the data used for this analysis. We have investigated the effect of multiple primary vertices on
### 3.2 Results

#### Figure 2: Distributions of dimuon invariant mass (data points) for tag muons paired with passing (left) and failing (right) probes in collision data. The mass is calculated using only the muon detector information for all probe muons. We do not use the silicon track momentum for the passing probes.

#### Table 1: Measured tracking efficiency values from tag and probe on data and simulation, after correcting for the effect of spurious muon-track matches. We show results for different pseudorapidity ranges, and for the combined result with $|\eta| < 2.4$.

<table>
<thead>
<tr>
<th>Region</th>
<th>Data Eff. (%)</th>
<th>Sim Eff. (%)</th>
<th>Data/Sim</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.0 \leq</td>
<td>\eta</td>
<td>&lt; 1.1$</td>
<td>$100.0^{+0.0}_{-0.3}$</td>
</tr>
<tr>
<td>$1.1 \leq</td>
<td>\eta</td>
<td>&lt; 1.6$</td>
<td>$99.2^{+0.8}_{-1.0}$</td>
</tr>
<tr>
<td>$1.6 \leq</td>
<td>\eta</td>
<td>&lt; 2.1$</td>
<td>$97.6^{+0.9}_{-1.0}$</td>
</tr>
<tr>
<td>$2.1 \leq</td>
<td>\eta</td>
<td>&lt; 2.4$</td>
<td>$98.5^{+1.5}_{-1.6}$</td>
</tr>
<tr>
<td>Combined</td>
<td>$98.8^{+0.5}_{-0.5}$</td>
<td>$99.2^{+0.1}_{-0.1}$</td>
<td>$0.996^{+0.005}_{-0.005}$</td>
</tr>
</tbody>
</table>
muon tracking efficiency by separating the $J/\psi$ data according to the number of reconstructed vertices. Figure 3 shows the resulting muon tracking efficiency as a function of the number of vertices, where no significant trend is observed.

4 Non-isolated Muons

The tracking efficiency measured using the tag-and-probe technique described in the previous section is valid for relatively isolated muons. However, it is also important to understand the tracking efficiency for charged particles in the more dense environment of jets, in particular for measurements that rely on the identification of jets containing bottom quarks ($b$ jets). In this section we describe a technique for measuring the tracking efficiency of muons in $b$ jets by matching standalone muons with tracks reconstructed in the inner tracker.

4.1 Method

We select events with at least one standalone muon satisfying $|\eta| < 2.4$, $p_T > 5\text{ GeV}/c$, more than ten hits, and $\chi^2/\text{ndf} < 10$ for the muon-segment fit. If there is more than one standalone muon passing this selection in an event, we take the one with the highest transverse momentum. We then require at least two jets reconstructed with a particle-flow algorithm [5] and having a minimum transverse momentum of $10\text{ GeV}/c$. One jet, referred to as the muon jet, is required to have its axis lie within a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ around the standalone muon direction, while the second jet, referred to as the $b$ jet, is required to be separated from the muon jet by an angular distance of at least $\Delta R = 1.5$, and to pass the standard $b$-jet tagging requirement referred to as Track-Counting High-Purity [6]. If there is more than one muon associated with the muon jet, we veto the event. For a data sample selected with a minimum-bias trigger and corresponding to an integrated luminosity of approximately $9\text{ nb}^{-1}$, we find 477 events satisfying all of the above criteria.

The predicted composition of muons in the selected sample is estimated from Monte Carlo simulation to be approximately 80% from $b$ and $c$ hadrons, and 20% from decays in flight of
4.2 Results

Figure 4: Distributions of $p_T^{\text{rel}}$ for standalone muons in data (points with errors) for events where the non-muon jet does (left) or does not (right) pass the $b$-tagging algorithm. The data are fit to templates for $b$ and non-$b$ jets derived from simulated samples.

light hadrons. Since muons from light-hadron decays could have a different tracking efficiency than muons from the decay of bottom and charm hadrons, it is critical to know the composition of the muon-jet sample in data, and the efficiency for reconstructing muons from light-hadron decays. We perform the measurement in two steps: first, we measure the composition of the $b$-tagged sample and a second sample where the $b$ jet fails the $b$-tag requirement, then we measure the tracking efficiency in each sample, from which we determine separately the tracking efficiencies for muons from decays of heavy-flavor and light-hadron jets.

We measure the fraction of muon jets from $b$-hadron decays in data by fitting the distribution ($p_T^{\text{rel}}$) of the muon momentum transverse to the jet direction, which is sensitive to the mass of the object that produced the muon jet. For the smaller sample of $b$-tagged jets, the resolution on $p_T^{\text{rel}}$ is not sufficient to separate the components for charm jets and jets coming from the fragmentation of light quarks and gluons, so we estimate the ratio of charm to light-hadron components from simulation and perform the fit using only two templates: one for bottom and one for non-bottom jets. For the larger sample of events that fail the $b$-tag requirement there is sufficient resolution to determine the relative fraction of the three components for $b$, $c$, and light-hadron jets. Figure 4 shows the resulting fit to $p_T^{\text{rel}}$ for data events that pass or fail the $b$-tag requirement, where we find $b$ fractions of $(71 \pm 11)$% and $(23.6 \pm 5.6)$%, respectively.

4.2 Results

We attempt to match the standalone muon to tracks in the muon jet that have at least two pixel hits, eight strip hits, and $\chi^2$/ndf < 10 for the track fit. A matching variable $a$ is defined as the distance between the extrapolated track and the position of the hit in the muon detector closest to the interaction region, divided by the distance from the muon hit to the center of CMS. The fractions of right and wrong track-muon associations are then determined by fitting the distribution of $a$ to templates derived from data. Finally, the efficiency is extracted by comparing the number of right associations to the total number of muons passing our selection criteria.

The right-match template is obtained from a data sample of highly pure global muons, using the distribution of $a$ calculated for the silicon track associated with the muon. The wrong-match template is obtained from the distribution of $a$ calculated for all tracks in the muon jet except than the one associated to the global muon. We have checked this method in simulated samples and have confirmed that the technique is unbiased.
Figure 5 shows the resulting fit to $\alpha$ on data that pass or fail the $b$-tag requirement, from which we derive uncorrected tracking efficiencies of $(91.4 \pm 5.6)\%$ and $(87.0 \pm 0.1)\%$, respectively. We correct these measured efficiencies for the presence of light-quark and gluon jets by solving a system of two simultaneous equations derived from the composition and efficiencies determined in the two independent samples, where we assume that the tracking efficiencies for muons from the decays of $b$ and $c$ hadrons are equal. From this calculation we find $\epsilon_{bc} = (93.2 \pm 5.3)\%$, where the uncertainty is statistical only. The method has been validated with a detailed check on simulated events, where the measured efficiency agrees with the true efficiency (96%) within 2.5%. The value measured in data is also in agreement with the true efficiency within its uncertainty.

5 Pion Tracking Efficiency

It is expected that tracking efficiency for hadrons will be different than that for muons due to decays and nuclear interactions with the material of CMS. Therefore, the efficiencies determined in the previous two sections cannot be applied to measurements involving the reconstruction of charged hadrons. In this section, we describe a method of estimating the relative tracking efficiency in data compared to simulation for pions in the momentum range 300 MeV/c to 1.5 GeV/c.

5.1 Method

The relative efficiency of reconstructing pion tracks in data and simulation can be determined by measuring the ratio of neutral charm-meson decays to final states of four or two charged particles. Specifically, we measure the production rate for $D^0 \rightarrow K^- \pi^+ \pi^- \pi^+$ ("$K3\pi$") relative to that for the two-body decay $D^0 \rightarrow K^- \pi^+$ ("$K\pi$") in both data and simulated samples. To increase the purity and provide a common production source, we reconstruct $D^0$ decays in the chain $D^{*+} \rightarrow D^0 \pi^+$. Assuming that the kinematic properties of the two decay modes are properly reproduced in the simulation, the ratio of efficiency-corrected signal yields,

$$R = \frac{N_{K3\pi}}{N_{K\pi}} \times \frac{\epsilon_{K\pi}}{\epsilon_{K3\pi}}$$

(4)
should be equal to the world-average ratio of branching fractions $\mathcal{R}(\text{PDG})$. The relative tracking efficiency for pions in data and simulation can then be estimated as

$$
\frac{\epsilon(\text{data})}{\epsilon(\text{MC})} = \sqrt{\frac{\mathcal{R}}{\mathcal{R}(\text{PDG})}},
$$

where $\mathcal{R}(\text{PDG}) = 2.08 \pm 0.05$ [7].

Reconstruction of $D^0$ candidates begins by selecting tracks with $p_T > 300\,\text{MeV}/c$ that have more than five hits, $\chi^2/\text{ndf} < 2.5$, and that are consistent with originating from the primary vertex [8] within 1 mm in the transverse plane, and 1 cm in $z$. Candidate $K\pi\pi$ and $K\pi$ decays are formed by vertexing four or two sets of tracks with zero total charge and requiring a vertex probability greater than 1%, and a positive decay length relative to the primary vertex. We then combine all $D^0$ candidates with an additional track (the “slow pion” $\pi_S$ from the $D^*$ decay) with $p_T > 250\,\text{MeV}/c$ and at least three hits, and require the mass difference $\Delta M$ between the $D^*$ and $D^0$ candidates to be less than 159 MeV/$c^2$. For the $K\pi$ decay we assign the kaon mass to the track whose charge is opposite to that of the slow pion, while for the $K3\pi$ candidates we select the charge assignment whose mass is closest to the $D^0$ mass. Finally, we retain all $D^*$ candidates with a $D^0$ invariant mass within 10 MeV/$c^2$ ($25\,\text{MeV}/c^2$) of the expected value for the $K3\pi$ ($K\pi$) mode, respectively, and require a minimum $p_T$ of 4.0 GeV/$c$ on all $D^*$ candidates.

### 5.2 Results

We determine yields in data and simulated samples using an unbinned maximum-likelihood fit to the $\Delta M$ distributions for events reconstructed in minimum-bias data corresponding to an integrated luminosity of approximately 0.47 nb$^{-1}$. For signal we use a single Gaussian function, while for background we use a phenomenological threshold function given by

$$
f = \left(1 - e^{-\left(\frac{\Delta M - \Delta m_0}{\Delta m_0}\right)}\right) \left(\frac{\Delta M}{\Delta m_0}\right)^p + p_2 \left(\frac{\Delta M}{\Delta m_0} - 1\right),
$$

where $\Delta m_0$ is the endpoint, taken to be the pion mass, and the shape parameters $p_i$, along with the mean and width of the Gaussian function for signal, are free to vary in the fit. Figure 6 shows fits of $\Delta M$ for the $K\pi$ and $K3\pi$ samples requiring $p_T > 5.5\,\text{GeV}/c$ for the $D^*$ candidates, which we use for our primary measurement result since it balances the two competing effects of background level and the kinematic range for the lowest momentum tracks. The fit returns 3363 $\pm$ 108 $K\pi$ and 2120 $\pm$ 114 $K3\pi$ signal events.

We measure the ratio $\mathcal{R}$ from the fitted signal yields and efficiencies obtained from simulation, and then use Eq. 5 to determine the ratio of tracking efficiencies in data and simulation as a function of the minimum $p_T$ on the $D^*$ candidates. Figure 7 shows the resulting ratio of tracking efficiencies, where we find values consistent with unity within the statistical uncertainties. We take as our measurement the $p_T > 5.5\,\text{GeV}/c$ bin, for which we find $\mathcal{R} = 2.11 \pm 0.14$ and $\epsilon(\text{data})/\epsilon(\text{MC}) = 1.007 \pm 0.034$, where the error is statistical only.

We have investigated several sources of systematic uncertainty on the measurement of $\mathcal{R}$, and its interpretation as the square of the ratio of tracking efficiency in data and simulation. These include dynamical effects related to the resonant substructure of the $K3\pi$ decay, the effect of nuclear interactions and possible bias from associated charge-asymmetric efficiencies (determined to be negligible with the current sample size), and uncertainty on the signal yields due to imperfect knowledge of the templates used to fit the $\Delta M$ distribution.

The $K3\pi$ decay proceeds through six distinct intermediate states, including vector and axial-vector resonances with various polarization states [9]. We estimate a systematic uncertainty by
Figure 6: Distribution of $\Delta M$ for $K\pi$ (left) and $K3\pi$ (right) events reconstructed in data requiring $p_T > 5.5$ GeV/c on the $D^*$ candidates.

Figure 7: The ratio of tracking efficiency in data and simulation as a function of the minimum $p_T$ for the $D^*$ candidate. The uncertainties are statistical only, and are correlated between bins due to the overlapping data samples.

varying the individual contributions of each of the six sub-decay modes within their branching fraction and efficiency uncertainties, and then take the sum in quadrature of the resulting variations in the total efficiency for the $K3\pi$ decay. We find a total systematic uncertainty of 1.3% on the ratio of tracking efficiencies in data and simulation from this source. We evaluate the systematic uncertainty on the $\Delta M$ templates by using two Gaussian functions to represent the signal peak, and by obtaining the background shape directly from the sideband region in the $D^0$ mass. From these variations we determine a systematic uncertainty of 0.5% due to uncertainty on the template shapes. The final result is $\epsilon(\text{data}) / \epsilon(\text{MC}) = 1.007 \pm 0.034 \pm 0.014 \pm 0.012$, where the first uncertainty is statistical, the second is systematic, and the third comes from the error on $R(\text{PDG})$. 
6 Conclusions

We have presented four methods for the determination of tracking efficiency in CMS. The track-embedding method demonstrates that the tracking efficiency for muons and pions is reproduced by the simulation within one percent. From the $J/\psi$ tag-and-probe technique we determine a systematic uncertainty on tracking efficiency of $1 - 2\%$ for isolated muons, depending on muon pseudorapidity, while the uncertainty for non-isolated muons is determined to be $5.3\%$ from a sample of heavy-flavour jets. The total uncertainty on pion tracking efficiency is estimated to be $3.9\%$ from a comparison of two-body and four-body $D^0$ decays in data and simulated samples. Except for the track-embedding method, the statistical precision of each result will improve with increasing integrated luminosity.

References


