

CMS tracking performance in Run 2 and early Run 3

Walaa Elmetenawee^{a,b,*} on behalf of the CMS collaboration

^a*INFN - Sezione di Bari, Via Giovanni Amendola, 173, 70126 Bari, Italy*

^b*Physics Department Faculty of Science, Helwan University, Ain Helwan 11795 Cairo, Egypt*

E-mail: walaa.elmetenawee@ba.infn.it

A precise and efficient tracking is one of the critical components of the CMS physics program as it impacts the ability to reconstruct the physics objects needed to understand proton-proton collisions at the LHC. The CMS detector has undergone extensive improvements in preparation for Run 3 of the LHC to operate efficiently at the increased luminosity and pileup. Significant algorithmic enhancements have been implemented to enhance the performance of the CMS tracking system. These enhancements concentrate on refining both track finding and selection processes. Performance measurements of the track reconstruction both in simulation and collision data will be presented. The performance is assessed using LHC Run 2 at $\sqrt{s} = 13$ TeV and early LHC Run 3 data at $\sqrt{s} = 13.6$ TeV.

The 32nd International Workshop on Vertex Detectors (VERTEX2023)

16-20 October 2023

Sestri Levante, Genova, Italy

*Speaker

1. Introduction

The track reconstruction in the CMS can be described as a 4-step procedure [1]. First, the algorithm constructs seeds. This step known as "seeding" provides the initial trajectory parameter estimates. These proto-candidates are progressed through the entire tracker, identifying compatible hits at each layer using the Combinatorial Kalman Filter (CKF) [2] algorithm and updating the track candidate and its parameters: this phase is referred to as "building." Following the combination of all associated hits, the tracks undergo a fitting process and are annotated with quality flags during the "fitting" and "selection" steps, respectively.

The iterative tracking algorithm employed in the CMS track reconstruction process, referred to as "Iterative Tracking", is based on the CKF. Initial iterations focus on identifying tracks that are relatively easier to detect, such as prompt tracks with higher transverse momentum (p_T). Hits associated with these reconstructed tracks are subsequently excluded from the hits set, effectively mitigating combinatorial complexity. This strategic removal of hits facilitates the reconstruction of tracks in more challenging kinematic regions. The complete set of iterations is summarized in Figure 1 [3].

2. Algorithmic improvements for Run 3

In preparation for the LHC Run 3, the CMS detector has undergone extensive improvements to operate efficiently at the increased luminosity and pileup. Notable hardware upgrades include the replacement of the innermost barrel pixel layer with new modules featuring enhanced front-end readout chips. Furthermore, significant algorithmic enhancements have been implemented to enhance the performance of the CMS tracking algorithm. These enhancements concentrate on refining both track finding and selection procedures. Additionally, efforts have been directed towards the development of tracking at High-Level Trigger (HLT) software, leveraging the utilization of heterogeneous computing architectures in order to improve the performance of the HLT both in terms of throughput and reconstructed object performance.

The track finding step experienced a significant development between the end of the LHC Run 2 and the beginning of Run 3. The CKF described above was the standard tracking algorithm employed by CMS in Run 1 and Run 2. In Run 3, a novel algorithm has been developed for track pattern recognition, known as mkFit [3]. This algorithm maximizes parallelization and vectorization in multi-core CPU architectures. Utilized within the CMS software for a subset of the CMS track reconstruction iterations, as illustrated in Figure 1, the mkFit algorithm maintains physics performance comparable to traditional CKF-based pattern recognition while enhancing the computational efficiency of CMS track reconstruction.

The tracking performance is evaluated in a simulated $t\bar{t}$ sample with superimposed pileup (PU) events by associating reconstructed tracks and simulated tracks. The number of generated PU events follows a uniform distribution ranging from 55 to 75. A reconstructed track is considered associated with a simulated particle if more than 75% of its hits originate from that particle [1]. If this criterion is not met, the reconstructed track is classified as a random hit combination and designated as a fake track. Simulations consider detector conditions without module failures, taking into account the residual radiation damage from Run 2 operations.

Iteration	Seeding	Target track
Initial	pixel quadruplets	prompt, high p_t
LowPTQuad	pixel quadruplets	prompt, low p_t
HighPTTriplet	pixel triplets	prompt, high p_t , recovery
LowPTTriplet	pixel triplets	prompt, low p_t , recovery
DetachedQuad	pixel quadruplets	displaced--
DetachedTriplet	pixel triplets	displaced--recovery
MixedTriplet	pixel+strip triplets	displaced-
PixelLess	inner strip triplets	displaced+
TopTec	outer strip triplets	displaced++
JetCore	pixel pairs in jets	high- p_t , jets
Muon inside-out	muon-tagged tracks	muon
Muon outside-in	standalone muon	muon

Figure 1: Iterations of the CMS track reconstruction sequence are presented in the table. The subset of iterations for which the mkFit algorithm is utilized is marked on the left. [3]

Simulated tracks coming from the signal (hard scattering) vertex with a transverse momentum (p_T) greater than 0.9 GeV and pseudorapidity ($|\eta|$) less than 3.0 are used in the performance computation. Tracking efficiency is defined as the fraction of simulated tracks associated with at least one reconstructed track. All reconstructed tracks coming from any vertex (including pileup vertices) are used in the fake rate computation. The tracking fake rate is defined as the fraction of reconstructed tracks not associated to any simulated particle.

The tracking fake rate is consistently lower when utilizing mkFit for track building in a subset of six tracking iterations (indicated in Figure 1), as illustrated in the left panel of Figure 2. Furthermore, employing mkFit leads to a substantial reduction in track-building time by approximately 3.5 times when considering the sum of time across these six iterations as highlighted in the right panel of Figure 2. Examining individual iterations where mkFit is applied, the reduction factor varies, ranging from about 2.7 to approximately 6.7.

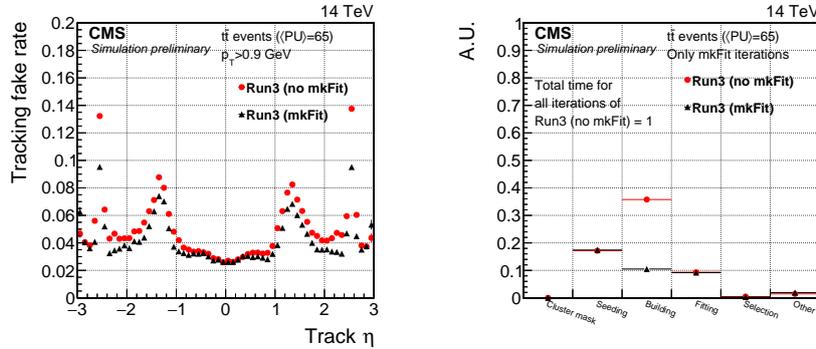


Figure 2: Tracking fake rate as a function of the track pseudorapidity η (left) and the tracking time as a function of the tracking steps, for the subset of six tracking iterations using mkFit for track building (right). [4]

After each iteration in the iterative tracking procedures for track reconstruction, a track classification step is implemented. This involves applying a multivariate analysis classifier and defining several selection criteria. If a track meets the high purity requirement, its hits are removed from the hit collection, thus simplifying the later iterations, and making the track classification an integral part of the reconstruction process. Tracks passing loose selections are also saved for physics analysis usage. The CMS experiment improved the track classification starting from a parametric selection used in Run 1, moving to a Boosted Decision Tree (BDT) in Run 2, and finally to a Deep Neural Network (DNN) in Run 3. The DNN was developed initially for pure CKF reconstruction, but

later for both mkFit and CKF tracks, which can have slight differences. These differences required training the DNN twice using two different sets of input tracks while keeping all the training settings unchanged. Reference [5] provides more information about the training procedures. In the current default tracking a DNN trained on mkFit tracks is used for the mkFit iterations, while a DNN trained on the CKF tracks is used for the CKF iterations.

The DNN performance is shown in the current default tracking (mkFit + CKF Run 3) and compared to the result of the Run 2 BDT on the same tracks. The physics results are shown after applying the high purity BDT or DNN selection to each iteration and after merging all the tracks from the iterations into one collection. The utilization of the DNN results in a lower tracking fake rate compared to the one obtained with the BDT, as illustrated in Figure 3. The most significant reductions in fake rates occur in the tracker endcaps ($|\eta| > 2$) and the barrel ($|\eta| < 1$). Additionally, a reduction in the fake rate is observed across all radii values, with an approximate 30% decrease.

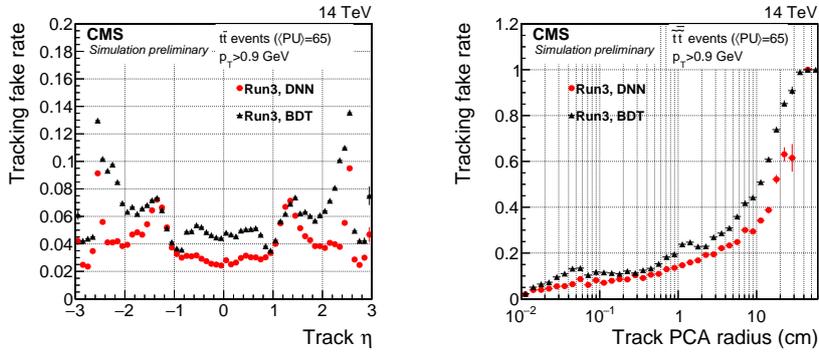


Figure 3: The tracking fake rates for tracks selected by DNN (red) and BDT (black) are compared in terms of track pseudorapidity η (left) and the radius of the track point of closest approach (PCA) to the beamline (right). [5]

The HLT of the CMS employs a streamlined version of full event reconstruction designed for fast processing. In the ongoing Run 3, the HLT utilizes a heterogeneous computing farm [6]. During Run 3, track reconstruction at the HLT is based on a single iteration of the CKF algorithm [2], utilizing hits recorded by pixel and strip detectors. This single iteration is seeded by pixel tracks reconstructed through the Patatrack algorithm [6, 7], which can be efficiently offloaded to Graphics Processing Units (GPUs). In order to be eligible as seeds, Patatrack pixel tracks must be constructed with a minimum of three pixel hits, possess a transverse momentum (p_T) exceeding 0.3 GeV, and be consistent with a leading pixel vertex.

Pixel vertices originating from primary interactions are reconstructed at the HLT during Run 3 through Patatrack pixel tracks featuring a minimum of four pixel hits and p_T exceeding 0.5 GeV. This reconstruction process involves clustering selected pixel tracks based on their z-coordinates at the point of closest approach to the center of the reconstructed ellipsoid approximating the luminous region. Subsequently, pixel vertices are arranged in descending order according to the summed p_T^2 of the associated pixel tracks. The vertex with the highest summed p_T^2 is labeled as the primary vertex (PV).

Figure 4 presents a comparative analysis of the HLT tracking performance in Run 2 and Run 3. In the left panel of Figure 4, tracking efficiency is presented, where an improvement of up to

20% is observed in the efficiency of Run 3 compared to Run 2. Moving to the middle panel of Figure 4, the tracking fake rate is showcased, demonstrating a notably smaller fake rate for Run 3 HLT tracking in the region with $0.9 < |\eta| < 2.1$, corresponding to the transition region between the barrel and the endcap detectors [7]. This is understood as in Run 3, pixel triplets are used to seed track reconstruction across the entire detector volume, providing improved performance in the transition regions compared to Run 2 HLT tracking, in which pixel triplets were confined to seeding the third iteration in regions around jet candidates identified from calorimeter information and tracks reconstructed in the preceding two iterations. The right panel of Figure 4 shows the track resolution d_{xy} , revealing a significant improvement in Run 3 performance across all regions compared to Run 2.

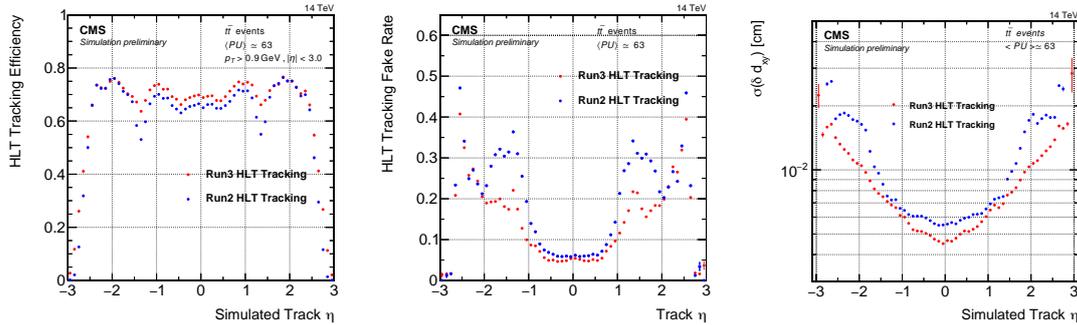


Figure 4: The plots illustrate the tracking efficiency (left), fake rate (middle), and track resolution (right) as a function of the track pseudorapidity (η) for both Run 2 HLT tracking (in blue) and Run 3 (in red). [7]

3. Tracking Performance using Tag and Probe Technique

The tracking performance can be characterized in the data by applying the tag-and-probe technique [1]. The efficiency is estimated using events with a Z boson decays into a pair of muons. The so-called "tag" muon is reconstructed using both the muon chambers and the tracker and identified by stringent requirements. The other muon ("probe"), instead, is reconstructed using only hits from the muon system and identified by looser requirements. Probes are then sorted into two categories, passing and failing, considering whether or not they can be matched to at least one track in a cone ($\Delta R < 0.3$) around the probe muon. The efficiency is calculated as the ratio between the passing probes and the total number of probes.

A tracking efficiency of about 99.9% for tracks associated with muons has been observed in the whole muon pseudorapidity acceptance through the full Run 2 data taking, as shown in Figure 5 (left). Figure 5 (middle) provides a comparative analysis of tracking efficiency across different running periods of Run 2. Notably, a period in 2016, denoted "2016 old APV," exhibited an inefficiency of up to $\sim 2\%$. This inefficiency was attributed to strip readout electronic saturation during high-intensity data-taking, a condition rectified in the remaining 2016 data, denoted as "new APV settings," following APV setting adjustments for fast recovery [8], resulting in a good agreement in efficiency with other data-taking periods (2017 and 2018 data).

Thanks to algorithmic improvements detailed in the previous section and the replacement of the innermost barrel pixel layer with new modules featuring enhanced front-end readout chips to

address issues identified in Run 2, the same excellent efficiency has been maintained for Run 3, as depicted in Figure 5 (right) for the barrel region.

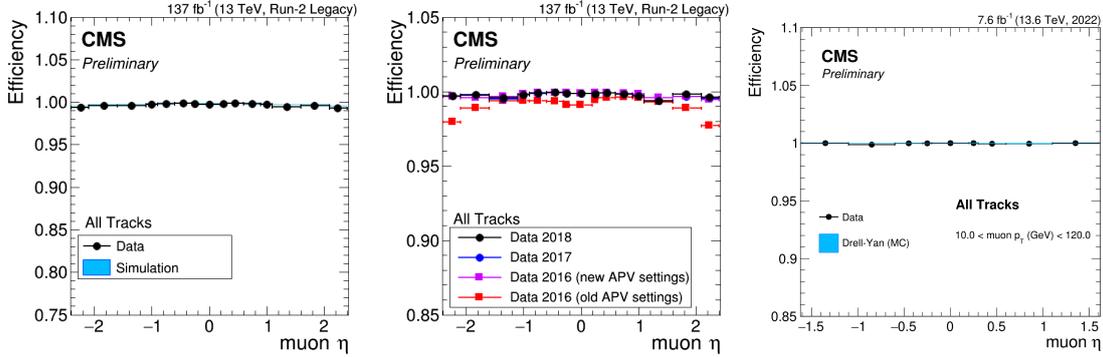


Figure 5: The tracking efficiency as a function of pseudorapidity η of the probe muon is shown for the full Run 2 Legacy [9] (left), the different eras of Run 2 data (middle), and for early Run 3 [10] (right).

4. Performance of tracking at High-level trigger (HLT)

The tracking efficiency and fake rate of the HLT, as measured in experimental data, are defined w.r.t. offline tracks generated through complete offline event reconstruction, which satisfy high-purity track quality criteria [1]. Both HLT and offline tracks are required to have a transverse momentum (p_T) exceeding 0.9 GeV, a transverse impact parameter relative to the primary vertex $|d_{xy}| < 2.5$ cm, and a longitudinal impact parameter relative to the primary vertex $|d_z| < 0.1$ cm. The matching between HLT and offline tracks is established by enforcing an angular distance criterion, requiring $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.002$.

The HLT tracking efficiency w.r.t. offline tracks is defined as the ratio between the number of HLT tracks matched to an offline track according to the requirement described above and the total number of offline tracks, while the fake rate is the ratio of HLT tracks that are not associated to any offline track and the total number of HLT tracks. Tracking efficiency (fake rate) relative to offline tracks, measured in data collected shortly before and after the LHC Technical Stop 1 (TS1), is represented as a function of the pseudorapidity of the offline track in Figure 6 left (right). Discrepancies in efficiency across the entire range arise from variations in efficiency in the Barrel Pixel Layer 1 (BPix L1). Several updates in the detector conditions took place during the TS1 of the LHC, including an increase in the Barrel Pixel Layer 1 (BPix L1) Bias High Voltage from 150 V to 300 V with a corresponding update of the pixel Cluster Position Estimator (CPE), as well as a new pixel detector gain calibration and a new tracker alignment which caused considerable improvement in the HLT tracking efficiency [11].

5. Conclusions

Significant algorithmic enhancements in the track finding and selection have been implemented to enhance the performance of the CMS tracking in preparation for the LHC Run 3, resulting in reduced tracking fake rates and total tracking CPU time compared to Run 2. The mentioned

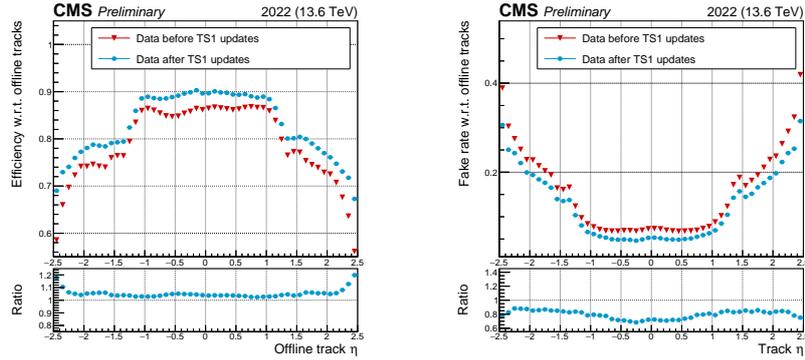


Figure 6: The HLT tracking efficiency (left) and fake rate (right) w.r.t. offline reconstruction is shown as a function of the offline track pseudorapidity η . [11].

improvements together with the installation of the replaced innermost pixel detector layer, effectively address the efficiency challenges, ensuring the CMS tracking system excels in the ongoing Run 3.

References

- [1] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, (2014) JINST 9 P10009.
- [2] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, Nucl.Instrum.Meth. A262 (1987) 444-450.
- [3] S. Berkman et al “Speeding up the CMS track reconstruction with a parallelized and vectorized Kalman-filter-based algorithm during the LHC Run 3”, arXiv:2304.05853.
- [4] CMS Collaboration, “Performance of Run 3 track reconstruction with the mkFit algorithm”, CMS-DP-2022-018.
- [5] CMS Collaboration, “Performance of the track selection DNN in Run 3”, CMS-DP-2023-009.
- [6] A. Bocci et al.,(2020) “Heterogeneous Reconstruction of Tracks and Primary Vertices With the CMS Pixel Tracker”, Frontiers in Big Data, 3, 601728.
- [7] CMS Collaboration, “Performance of Run-3 HLT Track Reconstruction”, CMS-DP-2022-014.
- [8] M. J. Frech et al., “Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker”, Nucl. Instrum. Meth. A466, 359 (2001).
- [9] W. Elmetenawee, “CMS track reconstruction performance during Run 2 and developments for Run 3”, <https://doi.org/10.22323/1.390.0733>.
- [10] CMS Collaboration, “CMS Tracking performance in Early Run-3 data using the tag-and-probe technique”, CMS-DP-2022-046.
- [11] CMS Collaboration, “Performance of Track Reconstruction at the CMS High-Level Trigger in 2022 data”, CMS-DP-2023-028.