Vertexing and Tracking Algorithms at High Pile-Up

Giuseppe Cerati[*] **on behalf of the CMS Collaboration**

University of California, San Diego

E-mail: giuseppe.cerati@cern.ch

The challenges of track reconstruction with the CMS detector in a high-occupancy environment are presented. Particular focus is put on developments implemented during the LHC shutdown, targeting a substantial reduction in computing time and fake track rate with no or little efficiency loss. New developments, targeting enhanced performance for specific object are also discussed. Resulting track and vertex reconstruction performance in different pile-up samples are presented.

---

[*]Speaker.

1. CMS Track and Vertex Reconstruction

At the heart of the CMS experiment [1] is the world largest all-silicon tracker, composed of a Pixel detector (66M channels, pixel size: 100x150 μm^2) and Strip detector (9.6M channels, 80-180 μm pitch, 10-20 cm long). Strips can be either single- or double-sided; double-sided modules are glued with a 100 mrad stereo angle and provide 3D position measurement in global coordinates.

CMS tracking [2] is based on Kalman Filter and can be divided into 4 main steps: seeding (initial proto-tracks made of 2 or 3 hits plus beam spot or vertex constraint), pattern recognition (search for track hits), fitting (estimation of track parameters from the set of hits) and selection (rejection of bad quality or fake tracks). This procedure is repeated iteratively and, at the end of each iteration, hits associated to high quality tracks ("High Purity") are masked so that the combinatorics is reduced for the next iteration. Seven iterations were used in Run1 reconstruction:

Table 1: Tracking iterations used in Run1 reconstruction. R=$\sqrt{x^2+y^2}$ is the particle production radius.

IterN	Name	Seeding	Target Track
0	Initial	pixel triplets	prompt, high p$_T$
1	LowPtTriplet	pixel triplets	prompt, low p$_T$
2	PixelPair	pixel pairs	recover high p$_T$
3	DetachedTriplet	pixel triplets	displaced, R$\lesssim$4 cm
4	MixedTriplet	pixel+strip triplets	displaced, R$\lesssim$7 cm
5	PixelLess	strip pairs (TIB-TID-TEC)	displaced, R$\lesssim$20 cm
6	TobTec	strip pairs (TOB-TEC)	displaced, R$\lesssim$60 cm

The CMS vertex reconstruction process [2] is initiated with a track clustering stage based on a Deterministic Annealing algorithm achieving a resolving power of the order of Δz$\sim$1 mm. Clusters are then processed with an Adaptive Vertex fit resulting in space coordinate resolutions of O(10 μm) both in x-y and z. Reconstructed vertices are finally sorted according to higher Σp$_T$^{2}, so that the first vertex identifies the hard interaction point.

2. Tracking Challenges at High Pile-Up

A realistic pile-up (PU) scenario for the beginning of LHC Run2 is the following [3]: start-up at 50 ns bunch crossing (BX) collecting up to $\sim$1 fb^{-1} with <PU>$\sim$30, then operate at 25 ns BX with <PU> increasing from $\sim$25 to $\sim$45 and collecting $\sim$15 fb^{-1}. Beyond LHC Run2, typical PU for PhaseI and PhaseII LHC upgrades are 70 and 140 respectively. Tracking is essentially a combinatorial problem, so high PU is a challenge because of the increase in tracker occupancy. Iterative tracking is only a partial solution since, even after hits used by tracks from all iterations are masked, the fraction of unassociated hits is still well above 50%. On top of this, several detector effects play an important role. First of all, when more than one track crosses a double-sided strip module, ambiguities may produce ghost hits; the rate of ghost hits increases with PU and in the innermost layer it becomes larger than true hits at <PU>=40. As a consequence, the effect of PU is dramatic on iterations seeded by pairs of strip matched hits (PixelLess and TobTec); on the contrary,

pixel triplet seeded steps are linear (Initial) or close to linear (LowPtTriplet, DetachedTriplet) with respect to PU (Fig. 1(c)). Also, with 25 ns BX, out of time charged particles increase the occupancy of the detector; the effect is much larger for strips than for pixels ($\sim$45% vs $\sim$5% larger occupancy). The pixel detector is affected by dynamic inefficiencies (mainly due to saturation of the chip readout buffer), which, in the first pixel barrel layer, reach the 2% level already below an instantaneous luminosity of $1 \cdot 10^{34}$ cm^{-2}s^{-1}.
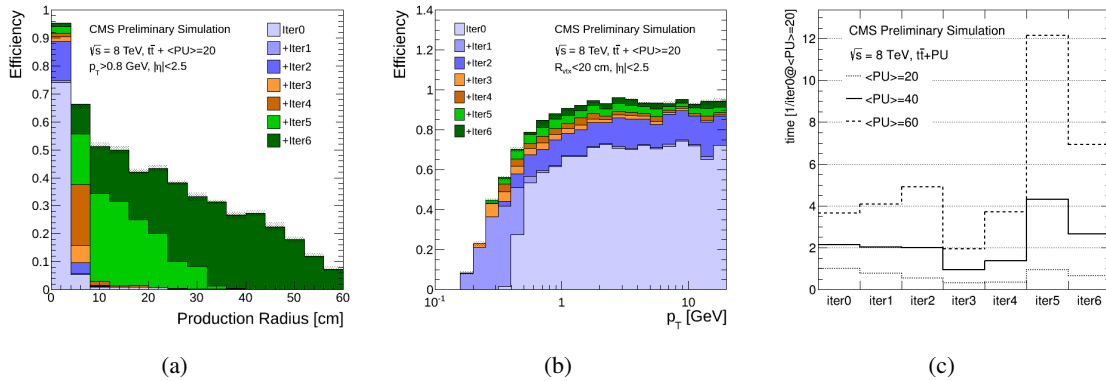

Figure 1: Performance of iterative tracking in Run1. From left to right: efficiency vs pseudorapidity, efficiency vs production radius and timing per iteration at different pile-up. Efficiency plots show the contribution from each iteration stacked in different colors.

3. Tracking Developments for Run2

During the first LHC long shutdown, CMS has developed a twofold strategy to face and overcome the challenge of tracking with high PU. First, the computing time needs to be highly reduced to fit into the time budget constrained by Tier0 reconstruction; this is achieved by designing a new seeding algorithm for strip-seeded steps, by reducing the combinatorics with new handles for rejecting hits and by optimizing both the code and the iterative tracking logic. Then, the physics performance can be further improved - both inclusively and for specific analysis objects like muons or jets. In fact, inclusive tracking has already very good performance and is the part that takes most of the reconstruction time; therefore, it can be improved only with lightweight developments like a new track selection. On the other hand, specific objects like muons or high p_T jets are relatively rare and contribute little to the overall processing time; therefore significant improvements can be made to the reconstruction of objects crucial for many analyses at a small price in terms of timing.

3.1 Timing-oriented Developments

The first priority of CMS tracking developments towards Run2 was the time reduction for the strip-seeded iterations, keeping a stable (or better) physics output. Despite the fact that most of the reconstruction time is spent during pattern recognition, seeding is the key for robust tracking against PU: first, a purer seed collection avoids wasting time trying to reconstruct fake tracks; second, a precise estimate of the seed parameters (and errors) optimizes the size of the hit-search window during the pattern recognition. Both goals can be achieved adding a third hit to seeds made of strip

pairs. Thus, a new seeding algorithm for PixelLess and TobTec iterations was developed. Main feature of the algorithm is a straight line fit of the 3 points in the RZ plane, where the hit position is dynamically re-estimated using the seed direction hypothesis for a higher precision; the resulting χ^2 is used as discriminant for the triplet selection. Beam spot constraints have been tightened as much as possible with no efficiency loss. The new algorithm is effective in rejecting half of the seeds, reconstructing approximately the same number of tracks (Fig. 2(a)-2(b)). Another issue mainly relevant for strip-seeded iterations is that, for 25 ns BX, 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. We thus introduced a cut on the cluster charge that can be applied with increasing strength at different stages of track reconstruction (upfront, seeding, pattern recognition) and that accounts both for sensor thickness and trajectory crossing angle. In Run2, stable selection efficiency will be ensured by the inclusion of the strip gain calibration in the "prompt Calibration loop" [4]. The cluster charge cut reduces the fake rate by ∼50% (Fig. 2(c)). Both new seeding and cluster charge cut reduce timing of PixelLess and TobTec iterations by a factor 2x (Fig. 3(a)).

Further time gain is obtained from a global optimization of the iterative tracking approach: iterations are re-ordered so that faster ones (in terms of time per track) are run first; redundancies are removed by reducing layer combinations for seeding and harmonizing the selection criteria. An overall 10% time gain is achieved, keeping stable physics performance. Significant speedup is also obtained from many code optimizations impacting both memory and CPU usage.

In order to benchmark timing and physics performance across releases and for different PU, we produced $t\bar{t}$ Monte-Carlo samples with realistic alignment and calibration conditions under various PU scenarios: <PU>=25, 40, 70, 140 at BX=25 ns and <PU>=25 at BX=50 ns. Figure 3(b) shows the time per event as a function of the instantaneous luminosity for the whole event reconstruction and for the tracking sequence only; time is evaluated both with the Run1 software release and with the current development release for Run2. Note that it was not possible to process the <PU>=140 sample with the Run1 release. A zoom into the region up to <PU>=70 (Fig. 3(c)) highlights that, in the new release, the tracking time reduction at <PU>=25, 40, 70 is respectively ∼2x, ∼3x, ∼4x and that the scaling is not far from linear in this range.

Special tracking sequences are widely used at High-Level Trigger (HLT) where, because of the increased beam energy and instantaneous luminosity of Run2, an even larger speedup is needed. We achieved a 4x time reduction with similar trigger performance at <PU>∼40 [5].

3.2 Physics-oriented Developments

As a result of the large speedup achieved, timing is well within the time budget for Run2 and special care could be devoted to improving physics performance. Main target for physics-oriented developments are specific reconstruction objects crucial for many CMS analyses.

The first example is represented by two iterations dedicated to muons. In 2012 data, a PU-dependent loss of muon reconstruction efficiency in the tracker was noticed. By adding an outside-in track reconstruction step seeded from the muon system, the muon-track efficiency is fully recovered (Fig. 4(a)). Also, an inside-out iteration that re-reconstructs muon-tagged tracks with looser requirements has been developed; it enhances the hit-collection efficiency (Fig. 4(b)), so that the muon momentum is more precise and the identification more robust.

The other object for which a dedicated iteration has been developed are high p_T jets; in fact, tracking in high p_T jets is crucial for efficient b- and τ-tagging and for resolving jet substructures. TeV-scale jets are a very dense environment characterized by small two-track separation, possibly causing merged clusters - where only one hit with badly-estimated position and uncertainty is reconstructed. In order to improve the track reconstruction performance in such environment, a dedicated, regional iteration has been developed; in a narrow cone around the direction of high p_T calorimetric jets merged pixel clusters are identified and split, then tracks are reconstructed testing in parallel a large number of candidates. The threshold on the jet p_T needs to be tuned to retain a smooth improvement in physics performance while limiting the increase in overall computing time. As expected, the new iteration improves the track reconstruction efficiency at small distance from the jet axis (Fig. 4(c)).

In order to improve the overall tracking performance, a multivariate track selection has been developed. Based on the BDTG algorithm implemented in the TMVA package [6] and using discriminating variables such as the number of found or lost hits, the track χ^2 or the p_T measurement uncertainty, it successfully increases the efficiency for low p_T and displaced tracks while at the same time reducing the fake rate in the forward region.

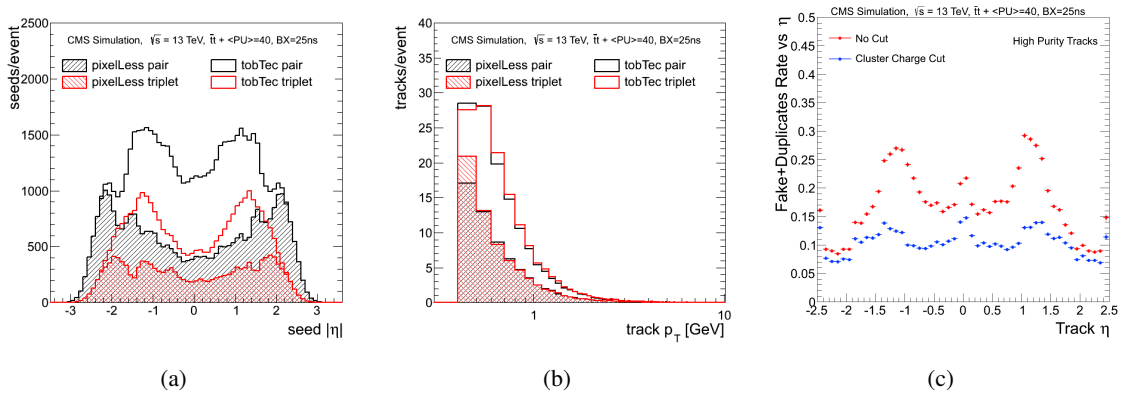| (a) | (b) | (c) |

Figure 2: Effect of the strip-triplet seeding and of the cluster charge cut. Left and center plots show the number of seeds and the number of tracks before and after using the new seeding algorithm. Right plot shows the reduction in fake rate due to the cluster charge cut.

4. Track and Vertex Reconstruction Performance vs Pile-Up

The impact of the presented developments is evaluated in terms of track and vertex reconstruction performance at different PU by comparing results obtained with the software release used for Run1 reconstruction and those obtained with the current Run2 candidate release.

Figure 5 shows plots of efficiency for high p_T prompt tracks vs eta, fake plus duplicate rate vs eta and efficiency for soft tracks vs production radius for three different PU configurations: Run1-like (<PU>=25, BX=50ns), Run2-like (<PU>=40, BX=25ns) and nominal conditions (Run1-like for Run1 release, Run2-like for current release). In all cases, the efficiency for prompt tracks is similar, while the efficiency for displaced tracks is slightly reduced in the current release. However, the reduction in terms of fake rate with the new release is up to a factor 2x for Run1-like and up to
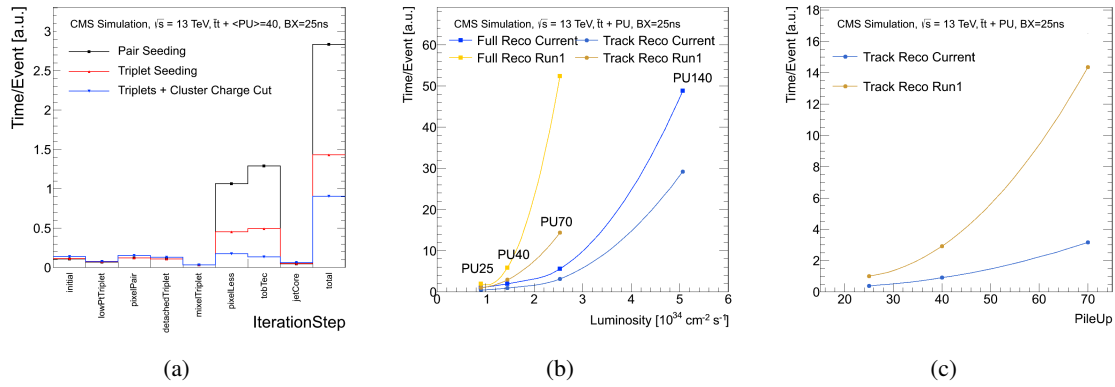
(a)   (b)   (c)

Figure 3: Timing measurements. Left plot shows the impact of the strip-triplet seeding and of cluster charge cut on PixelLess and TobTec iterations. Center plot shows the total and the track reconstruction time as a function of instantaneous luminosity for the Run1 and for the current Run2 candidate CMS software release. Right plot contains the same data but only the track reconstruction time is shown and the x axis is now zoomed and expressed in terms for PU. Y axes are normalized to Run1 track reconstruction at PU=25.
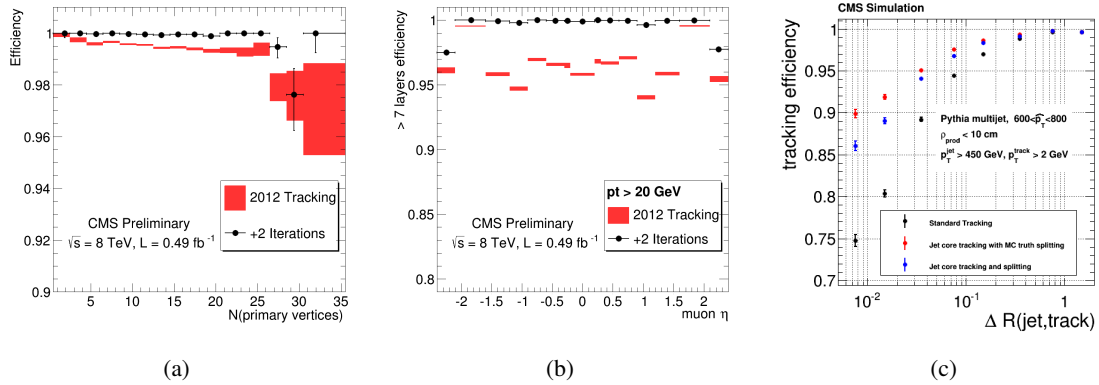

(a)   (b)   (c)

Figure 4: Results from object-specific iterations. Left plot shows the improvement in the muon reconstruction efficiency vs PU with the outside-in muon iteration; center plot shows the increase in hit collection efficiency vs muon η due to the inside-out muon iteration. Right plots shows the improved efficiency for track reconstruction in the core of high p_T jets.

6x for Run2-like PU. As a matter of fact, performance under nominal conditions are very similar and one can conclude that in Run2 CMS tracking performance will be close to those in Run1, but in a much more challenging environment.

Using the current Run2 candidate release the performance of vertex reconstruction at different PU are evaluated (Fig. 6). The number of reconstructed vertices as a function of PU shows a linear trend with a slope $\sim$0.7 up to PU$\sim$70, with a non-linear increase for higher values. On the other hand, when one requires the reconstructed vertex to be matched to a simulated one[1], the linear trend continues over the full range. These results are the effect of a faster than linear increase in fake rate and a mild linear decrease in efficiency as a function of PU. As a result, the efficiency for identifying the $t\bar{t}$ signal vertex as the one with highest Σp_T^2 is stable up to PU$\sim$70 and is definitely

---

[1]A reconstructed vertex matches a simulated if $\Delta z < 1$ mm and $\Delta z < 3\sigma_Z$.

sufficient for Run2; degraded performance at higher PU are due to degraded tracking performance with a detector designed for less extreme PU conditions.
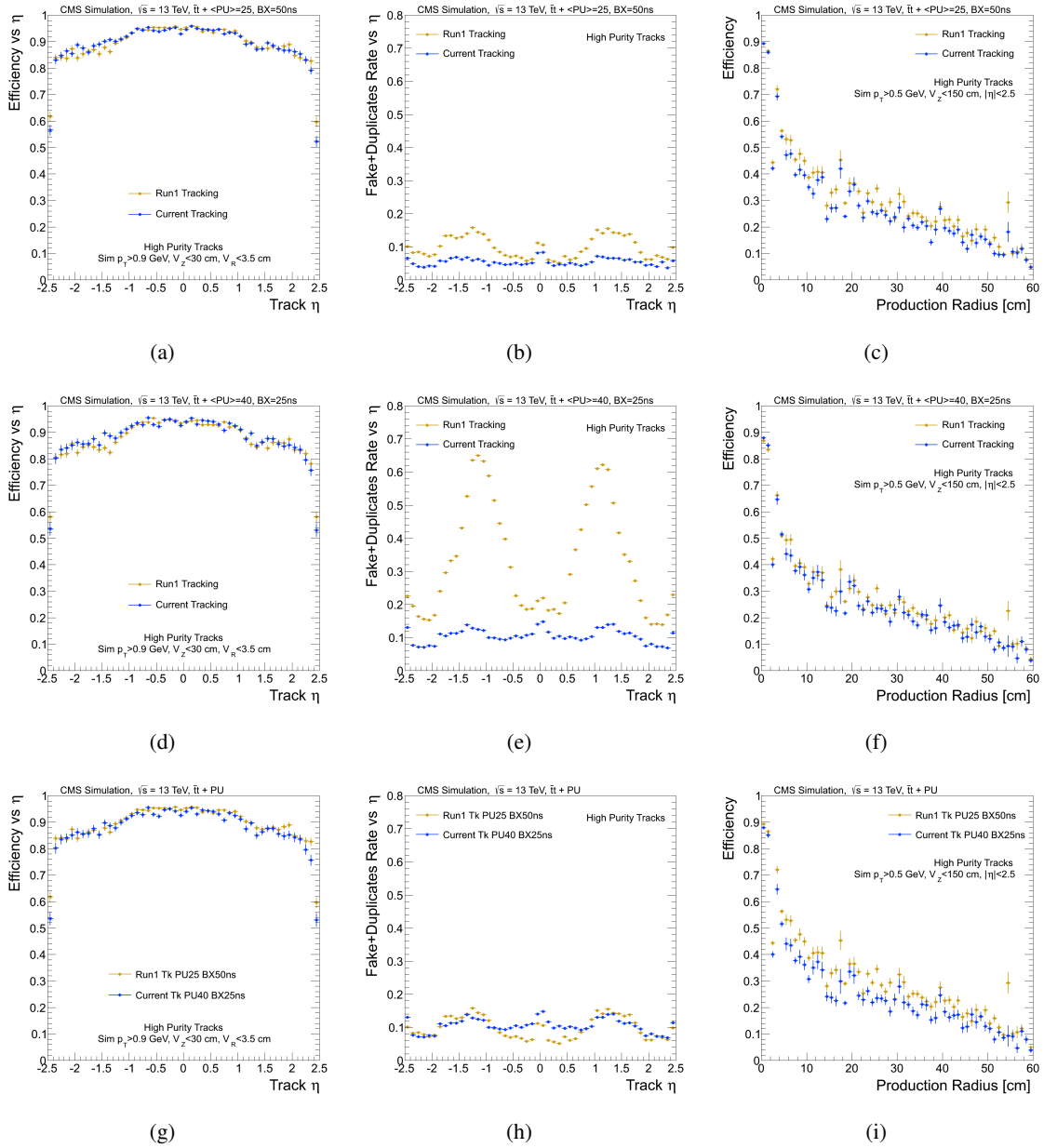

Figure 5: Tracking performance with Run1 and with Run2 candidate release. Each row shows efficiency vs η, fake rate vs η and efficiency vs production radius. From top to bottom, rows correspond to Run1-like PU, Run2-like PU and nominal PU conditions.

5. Outlook and Conclusions

High PU is a challenge for tracking due to occupancy and deterioration of detector performance. In preparation of the LHC Run2, CMS faced this challenge with a well defined strategy:
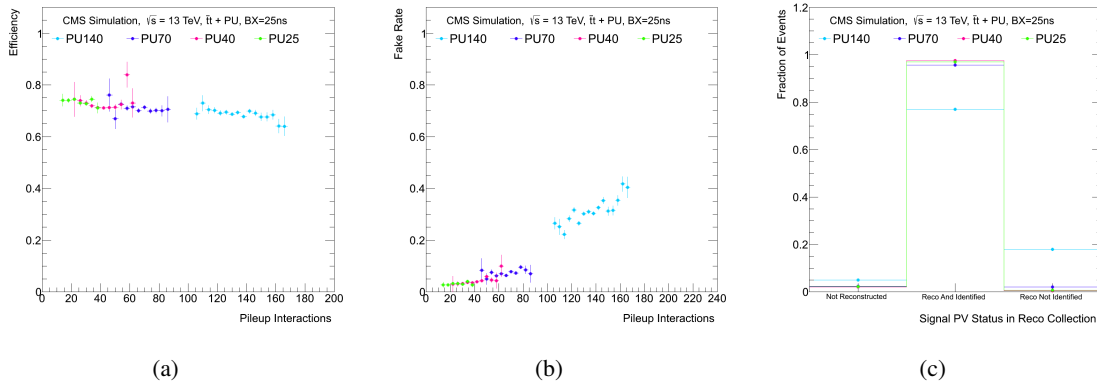
Figure 6: Vertex reconstruction performance for samples with different PU. Left and center plot show respectively the efficiency and fake rate as a function of the number of PU interactions. Right plot shows the fraction of $t\bar{t}$ events where the signal vertex is not reconstructed, reconstructed and identified or reconstructed and not identified.

first reduce combinatorics and then improve the performance of specific physics object. Results are excellent, with timing under control and Run2 performance comparable or better than Run1.

Beyond Run2, new challenges are coming for tracking at CMS. On one side, tracking algorithms need to be adapted to new detectors; in fact, a first implementation with upgrade geometries is already fully functional, both for the new pixel detector of PhaseI and the new tracker for PhaseII, including the extension of the forward pixel disks up to $|\eta|\sim4$. The PhaseII outer tracker features a fast and reduced readout for a L1 track trigger, whose implementation is under development. On the other side, the trend of new computing technologies goes in the direction of many-cores processors with large vector units; in order to efficiently exploit this kind of hardware, new algorithms need to be developed or old algorithms need to be reworked so that they are easily parallelizable and compliant with the SIMD (same-instruction-multiple-data) paradigm. As a first step in this direction the CMS tracking software is now thread safe and can be executed in multi-threaded processes.

References

[1] CMS collaboration, "The CMS experiment at the CERN LHC", 2008 JINST 3 S08004.

[2] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", 2014 JINST 9 P10009.

[3] J. Wenniger, "Overall Strategy for Run 2", LHC Performance Workshop, 22-25 September 2014 Chamonix - Les Aiglons.

[4] G. Cerminara, "Operation of the CMS detector with first collisions at 7 TeV at the LHC", PoS (ICHEP 2010) 485.

[5] M. Tosi, "Tracking at High Level Trigger in CMS", CMS CR-2014/253.

[6] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," PoS A CAT 040 (2007) [physics/0703039].