IL NUOVO CIMENTO **46 C** (2023) 99
DOI 10.1393/ncc/i2023-23099-4

Communications: SIF Congress 2022

# Offline primary vertex reconstruction for heterogeneous architectures

G. Pizzati(1)(2) for the CMS Collaboration

(1) *INFN, Sezione di Milano-Bicocca - Milano, Italy*
(2) *Dipartimento di Fisica, Università degli Studi di Milano-Bicocca - Milano, Italy*

received 31 January 2023

**Summary.** — The future development projects for the Large Hadron Collider will bring nominal luminosity increase, with the ultimate goal of reaching a peak luminosity of $5 \times 10^{34}$ cm$^{-2}$ s$^{-1}$. This would result in up to 200 simultaneous proton collisions (pileup), posing significant challenges for the CMS detector reconstruction. The CMS primary vertex (PV) reconstruction is a two-step procedure consisting of vertex finding and fitting. First, the Deterministic Annealing algorithm clusters tracks coming from the same interaction vertex. Secondly, an Adaptive Vertex Fit computes the best estimate of the vertex position in three or four dimensions. For High Luminosity LHC (HL-LHC) conditions, the reconstruction of PVs is expected to be extremely time expensive (up to 6% of reconstruction time). In this work, studies are presented on the rethinking and porting of PV reconstruction algorithms for heterogeneous architectures to exploit parallelization techniques to significantly reduce the processing time. We will show the results obtained focusing on computing and physics performance for HL-LHC conditions.

## 1. – Introduction

In high-energy physics experiments, the procedure to determine the position of the interaction point is known as *vertexing* and it is crucial to understand the dynamics of particle interactions. In CMS it is performed in two steps [1]: vertex finding and vertex fitting. This paper will present and discuss improvements for both procedures.

## 2. – Vertex Finding and Vertex Fitting at CMS

**2**˙1. *Vertex Finding with the Deterministic Annealing*. – Vertex finding consists in identifying the originating vertex for each track, and clustering them based on their impact point with respect to the beam spot. The cluster centers are used as the initial estimate of vertex positions.

To address this task, the CMS Collaboration implemented the Deterministic Annealing (DA) algorithm [2], which has proven to have excellent clustering performances. In summary, it is based on minimizing, at each iteration of an annealing schedule, a free energy function. Without any prior knowledge about the number of vertices, the final cluster configuration will be the one with the highest probability, making the DA algorithm independent from the pileup (PU) and thus a perfect candidate for vertex-finding for both LHC and HL-LHC conditions.

In the DA algorithm, each track is assigned to a vertex with a specific weight, *i.e.*, a probability of belonging to a specific vertex candidate.

Vertex finding through the Deterministic Annealing in a HL-LHC environment where PU interactions reach 200 on average is time expensive, taking up to 4% of the total event reconstruction time. The reason for such behavior is that the number of tracks to be clustered together is proportional to the number of PU vertices.

To speed up clustering, a wrapper of the DA was developed. The new algorithm, named *Deterministic Annealing in Blocks*, sorts the tracks based on their impact point z coordinate, splits them into sets of the same size (blocks), and runs the DA independently along blocks. Finally, it collects the resulting clusters, merging the ones that are too close (distance less than 1 mm). An overlap between blocks of tracks is introduced in order to avoid missing clusters of tracks belonging to adjacent blocks.

Two configurable parameters define the algorithm behavior: the block size, *i.e.*, the number of tracks in each block (set to 512 by default), and the overlap fraction between blocks (whose default value is 0.5).

**2**˙2. *Improving vertex fitting*. – The vertex finding procedure provides an estimate of tracks-to-vertex assignment as well as an initial rough guess of the vertex position, *i.e.*, the cluster center. *Vertex fitting* is the procedure that provides, starting from the clustered tracks' parameters, a more precise estimation of vertex parameters such as position, covariance matrix, and quality of fit ($\chi^2$).

To fulfill this task, the CMS Collaboration has adopted the *Adaptive Vertex Fitter* [3] which is based on a least squares method and includes an iterative estimate with an outlier rejection method in order to be a robust estimator.

To improve the task of vertex fitting, the least square estimator has been replaced by a weighted mean estimator.

The inputs of the algorithm are the tracks' covariance matrices and impact points at the beam spot. The estimate of the vertex position and error along a certain coordinate is computed through the weighted mean of the tracks' impact points using the reciprocal of the squared errors as weights.

To make the algorithm less sensitive to outliers, an iterative outlier rejection has been introduced in the weighted mean: only tracks whose impact point is compatible with the vertex are considered in the weighted mean iteration.

## 3. – Performances of algorithms

Computational and physics performances of the two novel algorithms were estimated with a $t\bar{t}$ simulated sample with 200 PU interactions superimposed per event. Four different configurations were used to test both two new algorithms in an independent way and in combination:

- Old Vertexing: the Deterministic Annealing combined with the Adaptive vertex fitter. The standard and current configuration in the CMS reconstruction software.

- New Clustering+Old Fitter: the combination of the *DA in blocks* with the Adaptive Vertex fitter in order to test the performances of the new clusterer alone.

- Old Clustering + New Estimator: the combination of the *Iterative Weighted Mean with Outlier rejection* with the default DA in order to test the performances of the new fitter alone.

- New vertexing: the combination of the two new algorithms developed.

Physics validation is based on the matching of a simulated vertex to a reconstructed one, performed in a geometrical fashion: given a simulated vertex, the 3D space around it is scanned, looking for a reconstructed vertex that is compatible with the simulated one.

The key metrics for physics performances are the efficiency and fake rate of vertices reconstruction, the resolution along the coordinates, and the correct identification and reconstruction of the hard scattering vertex.

The efficiency is defined as the ratio of the *number of simulated vertices that are matched with at least a reconstructed one* and the *total number of simulated vertices*. The fake rate is the ratio of the *number of reconstructed vertices that are not matched with any simulated one* and the *total number of reconstructed vertices*. The latter metrics are reported in fig. 1.

We define the resolution as the standard deviation of the distribution of the distances along a certain coordinate between a reconstructed vertex and its nearest simulated one.

In addition, to study the efficiency of reconstruction of the hard scattering vertex, we categorize the events in classes: 1) the primary vertex is not reconstructed, *i.e.*, no vertices match the simulated hard vertex, 2) the primary vertex is reconstructed but not tagged as the leading one, *i.e.*, the one with the highest $\sum p_T^2$, and 3) the primary vertex is both reconstructed and tagged as the leading vertex. Resolution and the latter metrics are reported in fig. 2.

Timings measurements on an Intel Skylake Gold processor are reported in table I. Ten batches with 2000 events each were used.

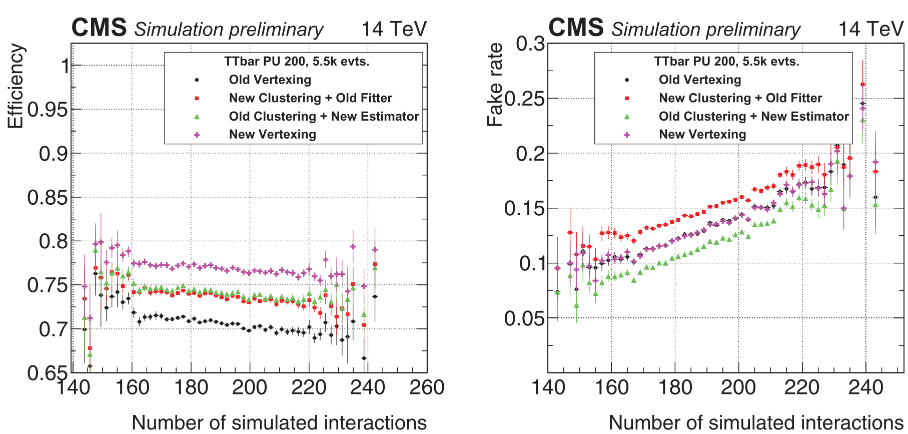


Fig. 1. – Efficiency on the left, and Fake Rate on the right, as a function of simulated pileup interactions.

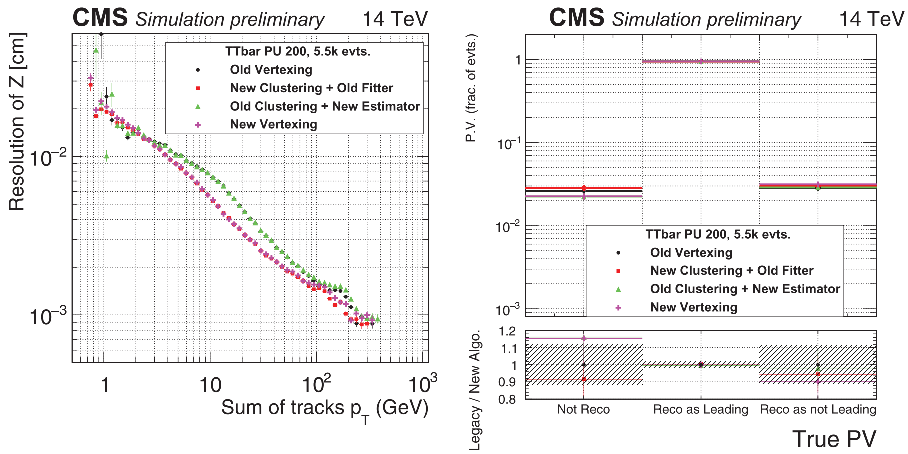


Fig. 2. – Resolution for the $z$ coordinate as a function of $\sum p_T^2$ of tracks on the left, hard scattering vertex reconstruction and identification performances on the right.

Table I. – *Timing performances of different PV configurations.*

| Algorithm label | PV reco. time (ms) |
|---|---|
| Old vertexing | $913.0 \pm 3.5$ |
| DA in blocks + Old fitter | $368.0 \pm 1.3$ |
| Old clustering + Weighted mean fitter | $512.6 \pm 3.3$ |
| New vertexing | $145.7 \pm 0.6$ |

## 4. – Conclusions

In conclusion, new vertexing algorithms were developed to speed up the vertexing task in HL-LHC conditions.

When used in conjunction, they allow reconstructing primary vertices 6.27 times faster than the current setup for $t\bar{t}$ events with PU 200. In addition, the physics performances of the new algorithms were tested, resulting unchanged or improved with respect to the current ones.

REFERENCES

[1] CMS Collaboration, *JINST*, **9** (2014) P10009.
[2] Chabanat E. and Estre N., *Deterministic Annealing for Vertex Finding at CMS*, *Computing in High Energy Physics and Nuclear Physics*, (CERN) 2005, pp. 287–290.
[3] Frühwirth R., Waltenberger W. and Vanlaer P., *J. Phys. G*, **34** (2007) 343.