Report Summer Students at Fermi National Accelerator Laboratory
CMS Phase II Track Trigger: Tower definition and optimization

Olmo Cerri

Dipartimento di Fisica "E. Fermi" - Università di Pisa
Scuola Normale Superiore di Pisa
Fermi National Accelerator Laboratory

October 11, 2016

Abstract

The purpose of this report is to document the work done by the author about the definition and optimization of trigger towers for CMS phase II upgrade.
Starting from the state of art at the beginning of the internship, a new definition of trigger tower is proposed and a study of its property is presented.

Contents

1 Introduction

One of the proposed solutions currently under study in Compact Muon Solenoid (CMS) collaboration [2] to reconstruct tracks at the first level trigger (L1) for the High Luminosity - Large Hadron Collider (HL-LHC) is based on the usage of Associative Memory [3] (AM) chips. The tracker information is first reduced to suppress low P_T tracks and sent to boards equipped with AM chips. Each AM compares the tracker information with pre-calculated expectations (pattern matching) in a very short time (order of a μs), therefore providing a solution to the challenging computational problem of pattern recognition in a very busy environment. Associated to fast track fit methods the AM approach should be able to fulfill the very demanding requirements of L1 tracking.

1.1 CMS tracking trigger

After the Phase-2 upgrades, the HL-LHC is expected to run at higher luminosity and higher energy and higher pile-up conditions compared with Run I or even Run III (after Phase-1 upgrades). The pile-up will drastically rise, up to 140 in average, potentially 200, i.e. much higher than the 50 expected in average for Run III and the 25 measured (with maximum value up to 35) in Run I.

The CMS Phase-2 detector will consequently have to keep the same performances than the current one but in a much more complex environment. This requirement is impossible to achieve at a Level-1 trigger with the current system requiring information from muon spectrometer and calorimeters only. A very important part of the HL-LHC physics program will be to look for very rare processes hidden in a huge amount of background. Implementing a powerful online event selection procedure will be essential to fully exploit the physics potential of upgraded CMS experiment where only a very limited fraction of the produced data can be recorded. In this context a brand new tracker associated to an intelligent tracking algorithm is required.

Due to significant damage and some inevitable performance losses at the end of Phase-1 running of CMS tracker and also to cope with new Phase-2 operating conditions, a new tracker is necessary. The current CMS Phase-2 tracker project is relaying on a barrel(layers)- endcaps(disks) geometry with a pixel inner detector and a outer tracker made of pixel-strip (PS) modules and two sided (2S) strips modules.
The outer tracker alone is considered for L1 track trigger. A sketch of one quarter of Phase-2 tracker

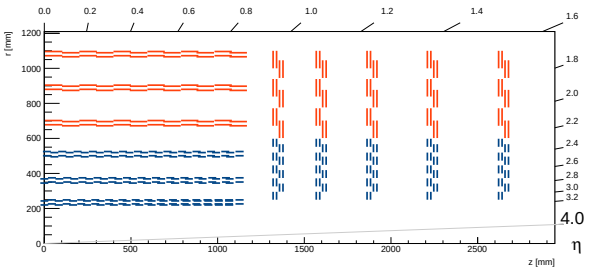

Figure 1.1: CMS phase II tracker

is shown on fig. 1.1. The outer tracker will provide data for both Level-1 trigger and global event processing in the High Level Trigger after positive decision from L1 trigger. In order to introduce tracking at Level-1 trigger level while coping with high pile-up conditions, a local reduction of the amount of data in the front-end is mandatory. This is done with outer tracker modules (Pixel-Strip and 2Strips), where the synchronized information from the two narrowly spaced sensors constituting one module can allow to reject low P_T tracks ($P_T <$ 2 GeV). We define *stubs* as pairs of clusters in the two sensors of a module within a predefined strips window. Sending stubs instead of clusters enable a first rough P_T cut at module level which can eliminate a large number of secondaries and also drastically reduces, by a factor 10-20, the amount of data to extract from the tracker at 40 MHz (50 Tbits/s of stubs).

Any tracking procedure can be defined in two steps. The first one, the pattern recognition (PR), isolates activated stubs potentially from a primary track while the second step, the track fit, determines more precisely the helix parameters of the track it finally reconstructs. To do this in high pile-up conditions, which implies about 20000 stubs in a single event, an intelligent tracking procedure, relying on Associative Memory (AM) chips for the first step and on a FPGA for the second one, will be detailed. This L1Tracking procedure will run at back-end level on dedicated trigger cards.

1.2 AM + FPGA approach

The complexity of the trajectory reconstruction and the high data processing require the development of custom hardware and intelligent tracking algorithms. A solution is to fulfill the L1 tracking with Associative Memory for the pattern recognition. AM have already been successfully used in high energy physics with the example of CDF experiment and will be implemented in the Fast Track project for ATLAS Phase-1 upgrade [4]. The benefits of using AM are:

- a fast processing
- a processing time independent of the number of hits/stubs.
- it is robust against pile-up
- it provides a clean environment to any track fit method: only stubs matched with track patterns will be considered in the track reconstruction step.

The AM based L1 tracking protocol begins with a data organizer (FPGA) receiving stubs at 40 MHz from the tracker and sending all of them to AM chips, containing pre-stored patterns banks, to perform the pattern recognition in one pass only. Once this operation done, matched patterns are recovered by the data organizer which finally send all necessary information, stubs and matched patterns for the event, to another FPGA performing the track fit. To cope with the huge amount of L1 data (50 Tbits/s), the tracker is divided into 48 trigger towers (TT): 8 sectors in (r-ϕ) plane and 6 in (r-z) plane. Each sector will consequently receive in average 200, and up to 500 stubs per event which represents around 400 to 600 Gbits/s per trigger tower.

2 Previous definition characterization

The first step of this work is to characterize the acceptance of the trigger towers as they were currently defined in tkLayout and identify possible problems. This study is conducted using TT27 as a model: this tower is a simple example of a central TT and was previously used for other studies.
The 4-dimensional parameter phase space for the track is slightly correlated between the two projection $\{q/P_T, \phi\}$ and $\{v_z, \eta\}$, therefore for this analysis is divided in the two subspace which are considered independent.

Starting from the $\{q/P_T, \phi\}$ projection, the shape of the current training sample for the tower in the parameter phase space is shown in the top of fig. 2.1. The training sample is made of single muon events with no pile up and is located in the file in the *l1upgrade* group folder in *cmsplc* called:
`stubs_tt27_300M_emu.root`.
From that sample only the tracks with $P_T > 2GeV$ and with with at least one stub in each barrel layer are considered and their MC truth parameters plotted.
That shape has to be compared with the shape in the phase space of the TT27 for which the sample is meant. In the bottom of fig. 2.1 is shown a plot of MC parameters of the tracks, coming from the same sample, with the same requirements as before and that have at least one stub per layer inside a module included in the TT27 definition. Looking at this comparison it is easy to understand that there is a mismatch between the training sample and the related TT definition: TT tracks phase-space region has some border in common with to the training sample which may be a clue for further region not included in the training sample. What is more, two missing corner are present.

These effects, caused by the fact that training sample are defined in terms of ϕ and η whereas TT in terms of ϕ^* and η^* (see B , lead to apparent inefficiency of the trigger and waste of memory and computational power.

To investigate the real structure of current TT definition a broad "global" sample of single muon event and no pile up has been generated: ϕ is allowed to be in all 2π range and $\eta \in [-2.4, 2.4]$. As it is possible to see from fig. 2.2, the actual definition of TT27 is indeed larger than its training sample coverage. As previously said, this mismatch is due to the different variables used in definition of TT and training sample.

Remembering that ϕ^* is defined as the ϕ coordinates of the particle position when it reaches a radius of $r = R^*$, that is $\phi^* = \phi(R^*)$, one can show that the boundaries in $q/P_T, \phi$ plane of a TT defined as an interval in ϕ^* must satisfy the equation:

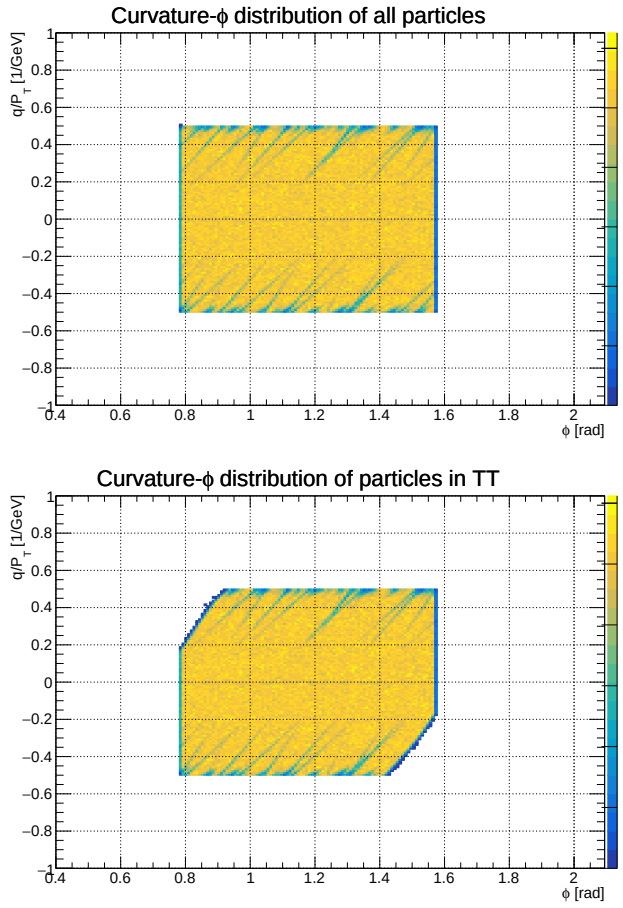

Figure 2.1: Comparison between the $q/P_T, \phi$ phase space of TT27 training sample and definition.
Top: All tracks inside training sample.
Bottom: All tracks in training sample with at least 6 stubs inside the TT definition.

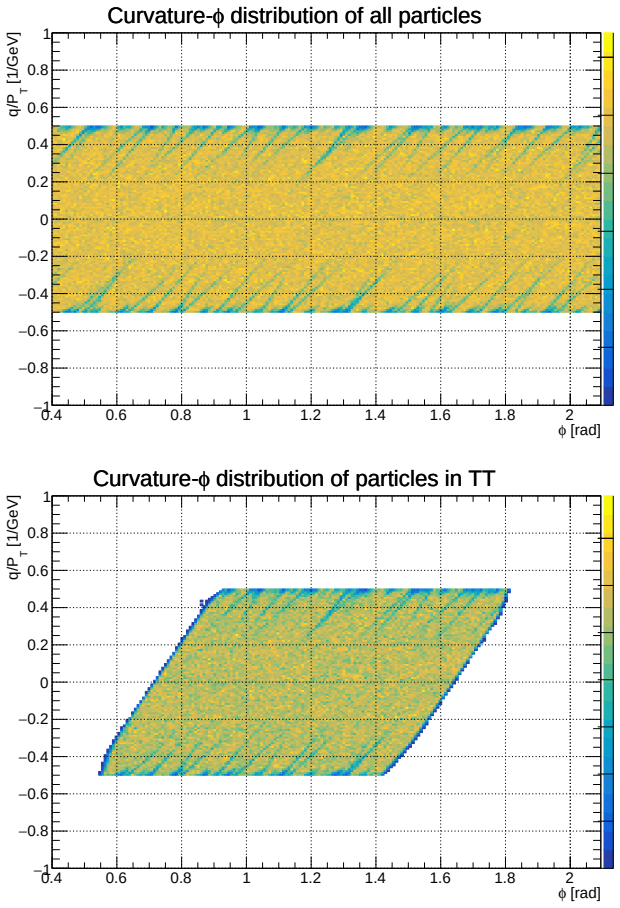

Figure 2.2: Comparison between the $q/P_T, \phi$ phase space of the global sample and TT27 definition.
Top: All tracks inside global sample.
Bottom: All tracks in global sample with at least 6 stubs inside the TT definition.

$$\frac{q}{P_T} = \frac{2\alpha}{R^*}\sin(\phi - \phi_e^{(1,2)}) \tag{2.1}$$

where $\alpha = \rho/P_T = 87.72$ cm/GeV for the CMS detector and $\phi_e^{(1,2)}$ are the two boundaries of TT in ϕ^*. A direct comparison of expected and effective border for TT27 and a resume of TT27 mismatch between definiton and training sample in shown in fig. 2.3. From this plot it is also possible to notice that there is an over coverage in phase space, this effect is due to the tracker modules finite dimensions.

The same analysis explained above for $\{q/P_T, \phi\}$ has been carried out for the $\{v_z, \eta\}$ projection. The results found are similar to the previous one and a summary plot is shown in fig. 2.4. In that picture it is possible to identify once again the mismatch due to usage of straight η border in the training sample generation. Here boundaries equations has not been explicitly evaluated but it is interesting to notice that TT27 definition border has a peculiar shape made of different slant region which is due to the detector geometry.
Further plot can be found in fig .A.2 and fig. A.1

2.1 Goal of the study

In light of all the above, the purpose of the study is to redefine all the trigger towers using the approach of the *star* parameters and perform optimization to reduce the towers overlap, overall number of modules and acceptance.
After that, for a given set of parameters, the list of modules to be assigned to each tower will be pro-

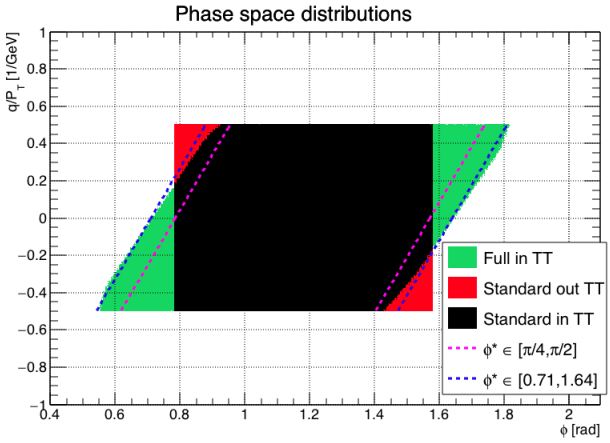

Figure 2.3: Direct comparison between parameters of tracks included in TT27 by definition and tracks in the TT27 training sample. In green the MC parameters from tracks in global sample which hit at least 6 different layer in TT27, in black and red tracks from TT27 training sample which hit at least (black) and less (red) than 6 layer in TT27.
Expected (purple) and effective (blue) border are evaluated using eq. 2.1.

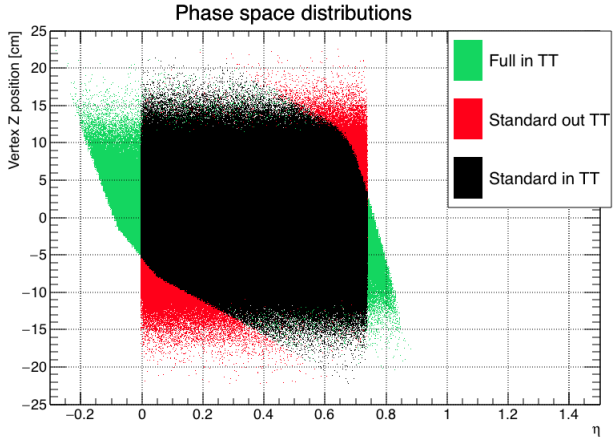

Figure 2.4: Direct comparison between parameters of tracks included in TT27 by definition and tracks in the TT27 training sample. In green the MC parameters from tracks in global sample which hit at least 6 different layer in TT27, in black and red tracks from TT27 training sample which hit at least (black) and less (red) than 6 layer in TT27.

duced, together with the physical boundaries of the tower in each layer of the tracker, in order to perform a full simulation of the track trigger system.
The final goal is to evaluate the goodness of the design through interesting parameters such as size of patterns bank, global acceptance or verify efficiency.

3 TT definition procedure

In order to define in a consistent way all the TT and to avoid bad definition, the first step is to divide the 4 dimensional parameters phase space into 48 non-overlapping regions. These regions, each one assigned to a single TT, have to fully cover the phase space.

3.1 Definition of phase-space

The phase space regions, to be assigned to a TT, are defined cutting the two projection $\{q/P_T, \phi^*\}$ and $\{v_z, \eta^*\}$ respectively into 8 and 6 slices. The 8 slices in ϕ^*, numbered from 1 to 8, have all the same width of $\Delta\phi^* = \pi/4$ and share their border with the adjacent slices. The TT with $N_\phi = 1$ starts from $\phi^* = 0$ and N_ϕ grows anti-clockwise.
The 6 slices in $\eta*$, numbered from 1 to 6, are equally spaces from $\eta^* = -2.2$ to $\eta^* = 2.2$ and have all the same width of $\Delta\eta^* = 2.2/3$. The TT with $N_\eta = 1$ includes all the tracks with $\eta^* \in [-2.2, -1.467]$ and N_η grows moving forward.
The trigger towers has a global numbering, from 0 to 47, given by:

$$N_{TT} = 8(N_\eta - 1) + N_\phi - 1$$

.

The shape of these regions is dependent from the choice of the values of R^*. When plotted on the two projection $\{q/P_T, \phi\}$ and $\{v_z, \eta\}$ the regions which are assigned to a TT keep the propriety of full coverage and no overlap but gain a *slant* aspect due to the variables change. An example is shown in fig .3.1.

Starting from the equations in appendix B, it is possible to verify that the border between 2 TT satisfy the following equations:

$$\begin{aligned} \frac{q}{P_T} &= \frac{2\alpha}{R^*_\phi} \sin\left(\phi_0 - \phi^*_{edge}\right) \\ \eta^* &= \text{asinh}\left[\frac{z_0 + 2\rho\sinh(\eta_0)\text{asin}\left(\frac{R^*_\eta}{2\rho}\right)}{R^*_\eta}\right] \end{aligned} \tag{3.1}$$

where R^*_η and R^*_ϕ are the values of R^* for the two projections.

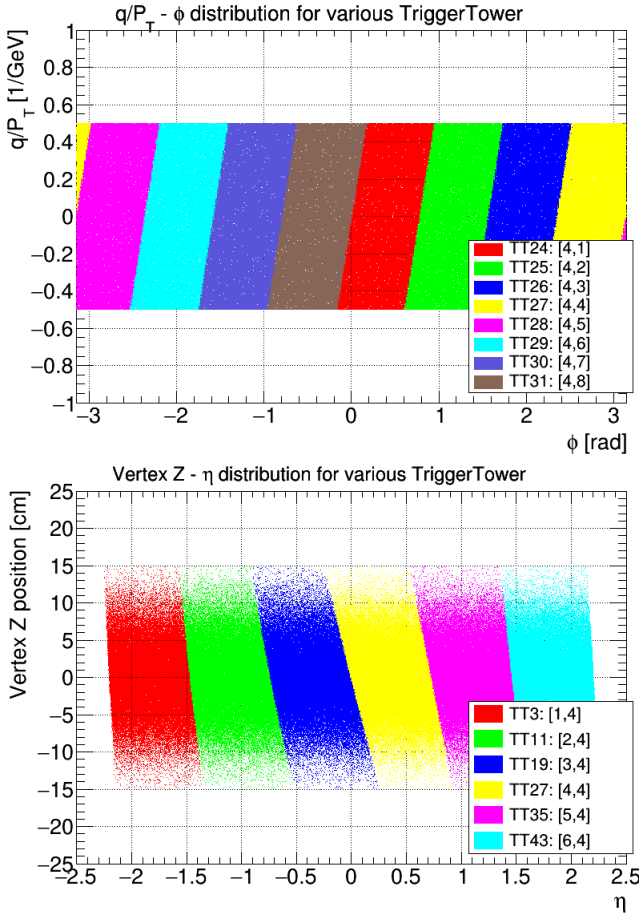

Figure 3.1: Region in tracks phase space assigned to different TT for $R^*_\eta = R^*_\phi = 59.89$ cm. Values at the primary vertex are meant for ϕ and η.
Top: Region in $\{q/P_T, \phi\}$ projection assigned to the towers in forward central zone ($N_\eta = 4$ and N_ϕ from 1 to 8).
Bottom: Region in $\{v_z, \eta\}$ projection assigned to the towers with $\phi^* \in \left[\frac{3}{4}\pi, \pi\right]$ (N_η from 1 to 6 and $N_\phi = 4$)

3.2 Definition of physical space

Once defined in phase space, in order to concretely define the physical dimensions of each tower, a full simulation approach is used.

The first step, given a TT, is to define a training sample made only of tracks which belong the the phase space region assigned to that TT. For this purpose, a wide generation of single muon events with no pile up is made and than divide into the 48 regions. The whole sample, made of about 50M events, contains tracks with $\phi \in [0, 2\pi]$ and $\eta \in [-2.4, 2.4]$. The momentum threshold is $P_T \geq 2$ GeV and Gaussian spread in primary vertex position (v_z) with $\sigma = 5$ cm is injected.

The second step is to go through all the tracks in a given training sample, that is a given TT, and keep a counters for each module of the tracker to record how many times every modules is hit. An

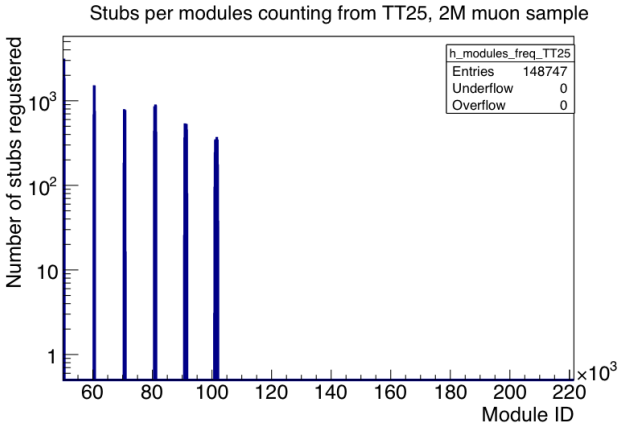

Figure 3.2: Number of stubs per modules produced by TT25 training tracks in a 2M single muon sample. Each bin is used as hit counter for a module.
Since TT25 is a central tower, only barrel modules ($modId < 110k$) has some stubs.

example of counts after this procedure is shown in fig. 3.2.

What is more, during the same procedure of going through all the tracks for a given TT, the physical boundaries of the region hit on each tracker layer is recorded.
This boundaries are essential for the definitions of super-strips. For each layer the boundary is defined by 4 number: maximum and minimum values of ϕ of the particles when hit that layer; the minimum and the maximum of z, for barrel, or r, for endcap, of the particles when hit that layer.
This allow to define a proper physical region where stubs must fall in to be sent to a specific TT.

4 New definition characterization

The above procedure leads to a full list of TT definition in both physical and parameter space. This procedure can be repeated for several configuration and parameters choice to pursue a comparison between different scenarios and look for the optimal setup.

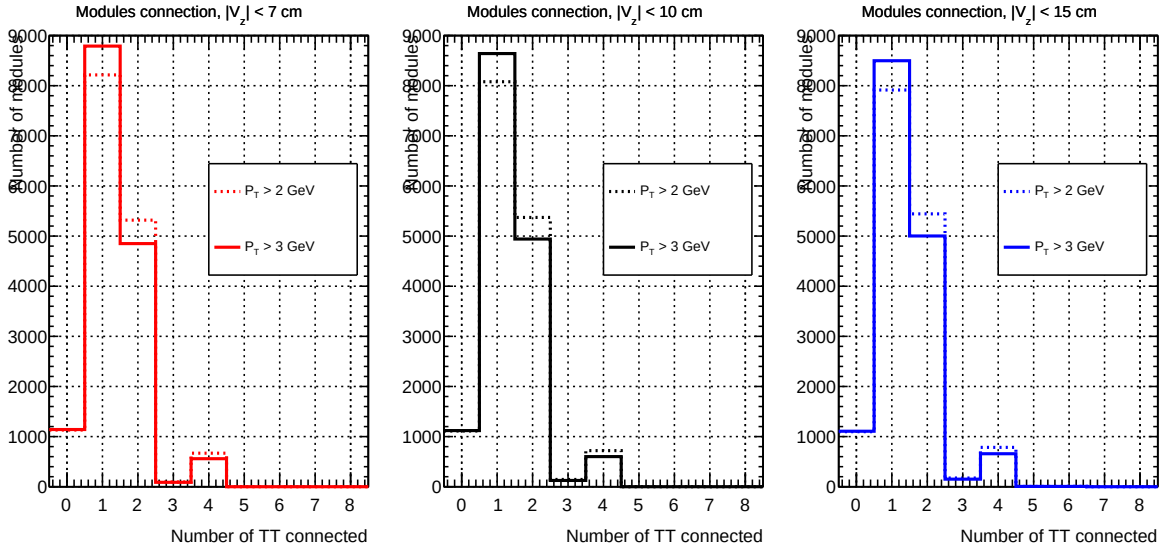

Figure 4.1: Distribution of the number of TT connected to a single module for various scenario of P_T and v_z acceptance. It is interesting to notice that the differences are of the order of few percent.

4.1 Module fan-out

The module fan-out distribution is presented in fig. 4.1 for 6 different configurations: three different width for the acceptance of primary vertex z coordinate (v_z) and, for each of them, two different cut on the minimum P_T. From that figure it is possible to see that the difference between the two different P_T configuration is of the order or few per cent and that the overall shape is almost independent from the v_z acceptance.
All configurations have modules fan-out equal or smaller than 6 and the large majority of modules connected to one or two TT.

Further studies has been done to understand what is the physical cause of each fan-out possibility:

- 0 connections: The TT definition in η reduce the acceptance and cut away tracks which are near the beam. For this reason there are several modules at high η that are not hit by tracks inside any training sample and so have no connection.

- 1 connection: Standard situation for a module in the core of a TT.

- 2 connections: Modules near the border of two TT may need to be connected to both of them because they are hit by tracks belonging to different TT. Both η and ϕ borders contribute to this effect.

- 3 connections: This is a rare situation, less the 100 modules require this fan-out and, usually, need to be connected to 2 adjacent TT in ϕ and one forward in η. For example the module 50028 has 3 connections (TT8:[2,1] - TT16:[3,1] - TT23[3,8]): two of the are adjacent tower in the central zone while TT8 is in the forward zone but in the same ϕ position of TT16.
  This kind of effect is due to the tracks bending: even if a track fall in to the intersection of two near ϕ region, it can move to the core of one of them while propagating to forward zone.

- 4 connections: Non common, about 600 modules need this fan-out and are all located on the edge which is in common to 4 TT. This is due to the coincidence of both η and ϕ border.

- 5 connections: Same as 6 connections but with the bending effect described for 3 connections which move the track in the core of one of the most forward TT.

- 6 connections: Very rare (about 10 modules) and only involving modules in layer 5. The modules with fan-out equal to 6 belongs to 3

consecutive TT in η and lay near the $\phi border$. For example the module 50029 has 6 connections (TT8:[2,1] - TT15:[2,8] - TT16:[3,1] - TT23[3,8] - TT24[4,1] - TT31[4,8]).

An example of tracks which produce the necessity of 6-time shared modules is show in fig. 4.2.

Purging 6-time shared modules - First of all it is important to said that 6-time shared modules are present only if v_z acceptance is larger than 10 cm and even in that case they are very rare.

In any case, assuming, as working hypothesis, that the maximum fan-out for a module is 4 a procedure to reduce the fan-out request has been applied.
For each module the connections are reduced to 4 removing that module from the modules list of the TT in which that module has the lower number of stubs recorded.

E.g. the module 50029, in a sample of 2M events, has been hit by 2 tracks belonging to TT8, 3 to TT15 and more than 5 tracks belonging to TT16, TT23, TT24 and TT31. During the purging procedure this module will be removed from the modules list of TT8 and TT15.

It is important to notice that the purging affect the efficiency for 6 out of 6 but does not affect the efficiency of 5 out of 6 (if you neglect modules detection efficiency)

Fan-out summary - To conclude in fig. 4.3 is reported the fan-out for different scenarios of v_z acceptance. It is important to notice that the differences are very small, order of few percent, and seems clear that enlarging the acceptance region from 7 cm to 15 cm is worth.

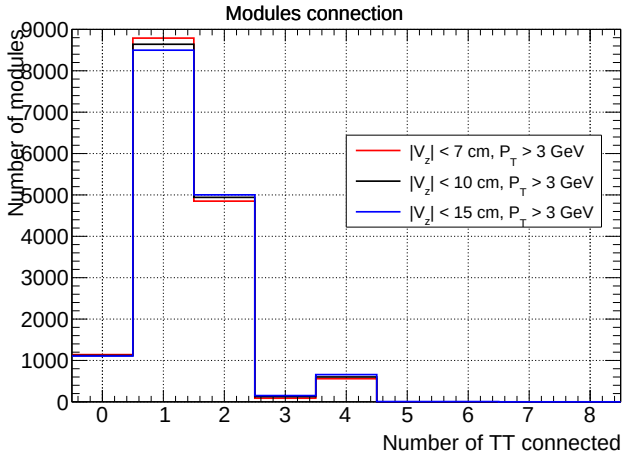

Figure 4.3: COMPARISON $P_T = 3GeV$

4.2 Tower connections

A different important factor to take care is the total number of modules connected to a single TT. This number, which vary of few percent from 2 GeV to 3 GeV boundary on lowest reco track P_T (fig. A.3), is of the order of 500.

A fair comparison between old and new TT definition can be found in fig. 4.4, the the old TT number has been changed to match the new one.

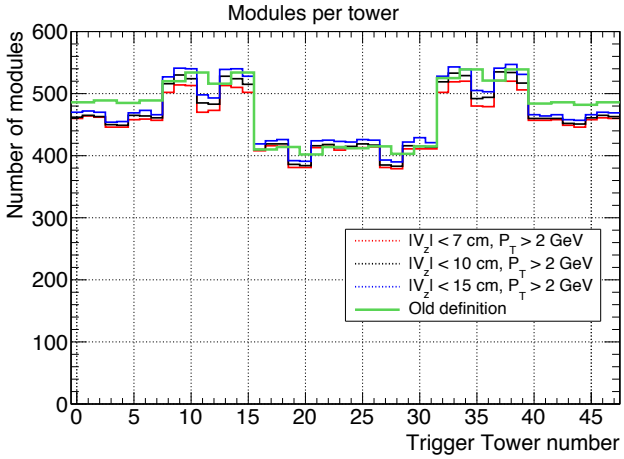

Figure 4.4: Comparison between the old and the new number of modules distribution.

It is clear that the new configuration is different from the old one but the number of modules connected to a TT has not a constant trend: some towers requires more connection than before, some others less. These differences are due to the different procedure of definition.

4.3 Output files

All the above procedure is meant to develop a consistent way of defining the TT, given some input parameters. First of all a 4π training sample which cover the whole detector must be provided to perform the simulation. On top of it, it is also needed to fix the values of the TT definition: v_z and P_T acceptance, R^*_η and R^*_ϕ.
Given the above infos, at the end of the definition procedure 2 files are produced as output:

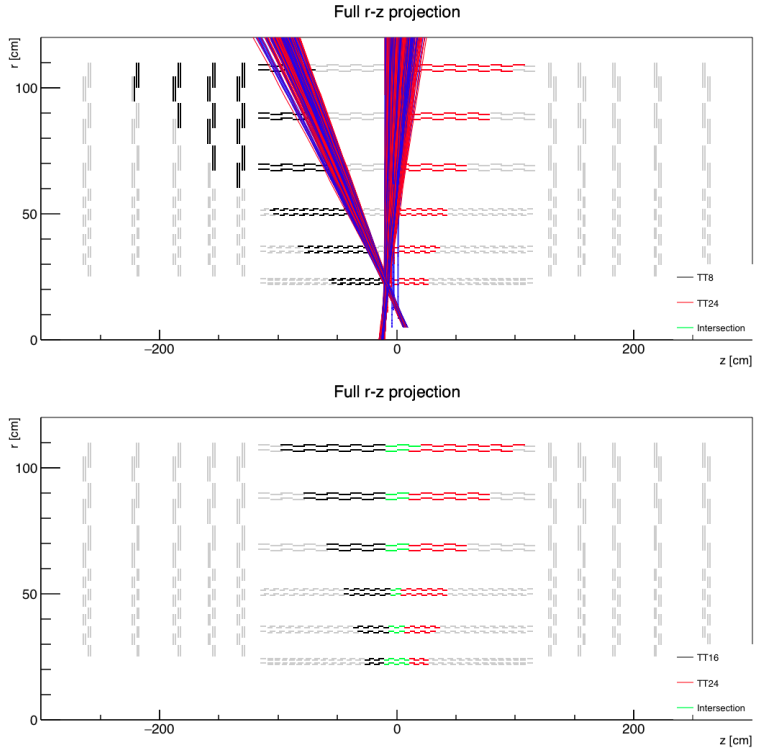

Figure 4.2: Top: Tracks with 10 *cm* $\leq |v_z| \leq$ 15 *cm which has one stub in module 50029 (6-time shared ones). It is possible to see that tracks with* $v_z < 0$ *and* $\eta > 0$ *eventually fall in one of the central forward tower (TT24) whereas tracks with* $v_z > 0$ *and positive* $\eta < 0$ *may belongs to the intermediate backward tower (TT8).*
Bottom: r-z projection of modules included in a central backward tower (TT16), central forward tower (TT24) and their intersection.

- Modules list: the list of all the modules belonging to each TT is produced. As standard output a *.csv* file is produced with the modules list for all the TT
- Boundaries list: a list with the boundaries in physical coordinates on each layer for all the TT. For barrel layers (5 to 10) the boundaries are in ϕ and z whereas for endcap disks the boundaries are in ϕ and r.
  These boundaries, used as extremes for the definitions of super strips, are obtained simply recording the extreme values in a specif coordinate of the particles inside the phase space assigned to a TT.

5 R* optimization

One of the goal of the work is to optimize the TT definition. There are several ways of doing it and two of them are presented in the following.
Since the $\{q/P_T, \phi^*\}$ and $\{v_z, \eta^*\}$ projection are not correlated the possibility of having different values of R^* for the 2 is explored. What is more the optimization analysis is performed separately for the 3 types of TT: central (barrel), intermediate and forward (endcap). Trying to optimize separately for each TT is useless because of the ϕ symmetry.
It is also necessary to keep in mind that, while the difference between R^*_ϕ and R^*_η imply no complications, different values of R^*_η for the three different region requires to take care of boundaries matching.

Geometrical consideration of minimizing the overlap area may lead to an optimal value of $R_{det}/\sqrt{2}$ where $R_{det} = 110$ cm is the radius of the last layer of tracker.

5.1 Number of modules

In this method the task is to minimize the number of connections needed to the tracking trigger system. The crude bottom line is to reduce the number of physical connections (optical fibers): this approach is worth if one of the bottleneck is the hardware.

To perform the study the following steps have been used:

- Fix a given value of R^* in one dimension (ϕ or η)
- Decide some values of the R^* of the other variable and for each value:
  1. Define all 48 TT af the detector and produce the modules list for each of them
  2. Sum the number of modules connected to all trigger towers inside each one of the 3 regions.
     For example to get the total number of connection in the endcap the number of modules of tower from 0 to 7 and from 40 to 47 has been summed up
- Record the total number of connections in the three regions and plot them against the value of the R^* scanned.

The results from this procedure (fig. 5.1) shows that there is a real gain moving towards higher values of R^* in both the projections. Both the scan in R^*_ϕ and R^*_η shows a plateau near the minimum but their behavior for extreme values is different.
The number of connections as a function of R^*_phi rise again after the minimum and may be not well defined for softer partcile which does not reach at all that value of r because of the bending.
Same is true for the scan in R^*_η that it also diverge near 0 because tracks inside a given η^* region may get across a increasing region of the detector.

Two more thing must be noticed: first of all, almost no dependence of these curves from v_z acceptance has been observed (more plots in the talks); second, this kind of optimization favor high values of R^*.
This is true because even if you increase R^* and, as consequence increase the size of the region on the inner layer included in the TT, the number of modules from the inner layer does not change unless you get out of one module which is relatively large wtr to the width of the TT region. On the other hand the outer layer region is much larger than one module and moving the value of R^* you can strongly change the number of modules included form the last layer.

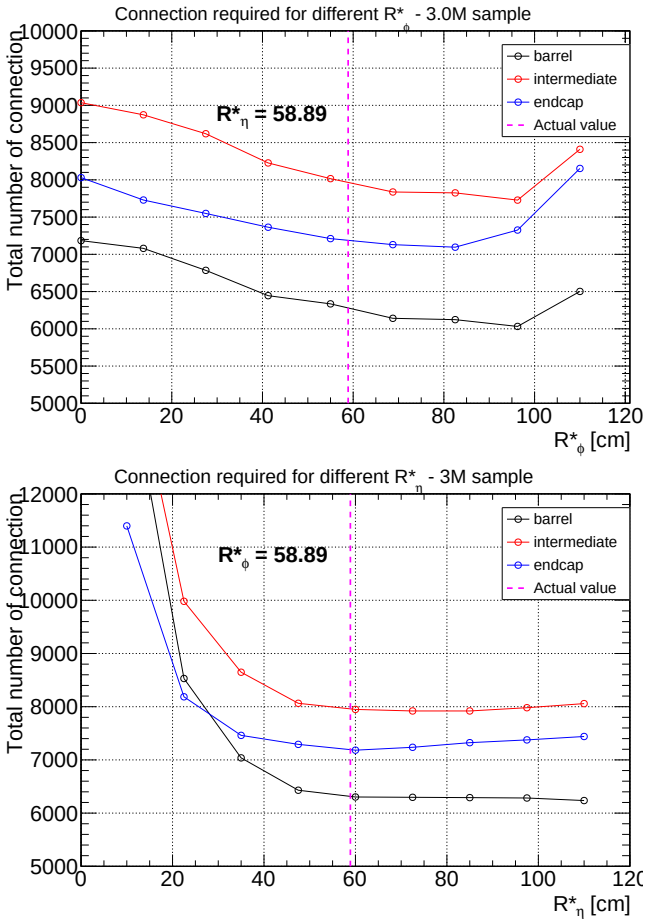

Figure 5.1: R^*_ϕ and R^*_η optimization scan for a TT definition with $|v_z| < 10$ cm and $P_T > 3$ GeV. In both cases the R^* on the other projection has been fiexed to a value of 58.89 cm.
For each of the three zones of the tracker the number of connection for each tower is summed together as explained in the text.

As conclusion this approach minimize connections but prefer to connect modules from the inner layer.

5.2 Weighted sum

This study has been performed to solve the same problem as the previous one but from a different point of view.
Indeed simply optimizing the number of modules does not takes in account the flux of information that each modules produce, point that can be crucial if the bottleneck of the process is software time to threat incoming data.

For example, pile up events, which are the most common ones, produce particles, so stubs, more of-

ten focalized near the beam pipe and that are more likely to overcrowd inner and forward modules. For this reason in this study a procedure similar to the previous one has been followed with the exception that during the sum (point 2 of the second bullet) the weights of each module included is summed instead the simply number of modules (the previous method is equivalent to give to all the modules a weight of 1).
The weight for a module is set as the average number of stubs in a events with PU140. This minimization procedure, that is equivalent to minimize the number of stubs that a tower need to process, disfavor inner and forward modules and therefore high values of R^*. From fig. 5.2, where the weight

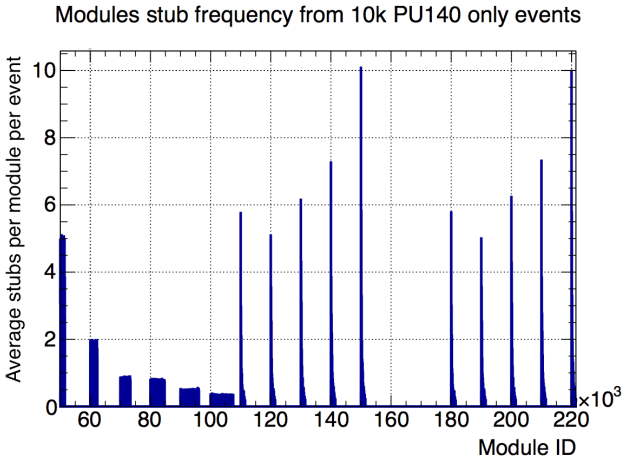

Figure 5.2: Average stubs produced in a pile up only events with 140 interactions. The bin entry for each module has been used as weight.

distribution is shown, it is clear that forward and inner layer have more stubs than layer from 6 to 10. It is also interesting to notice the particular shape of weight distribution inside one of the forward layer (fig. 5.3). Indeed, the module numbering is such that the first 2 numbers identify the layer, the second 2 the ring(endcap) or the ladder (barrel) and the remaining one the module in anti clockwise order. This imply that particular shape and makes clear the distinction between pixel-strip and strip-strip modules.

The results of the scan in R^*_ϕ and R^*_η, shown in fig. 5.4, are in agreement with expectations.
The optimal values are a bit smaller, especially in R^*_η, if compared with the ones from the previous study and minima are slightly deeper.

To conclude, this second approach in complementary to the previous one and the optimal choice, which should be a compromise between the two, has to be taken in future.

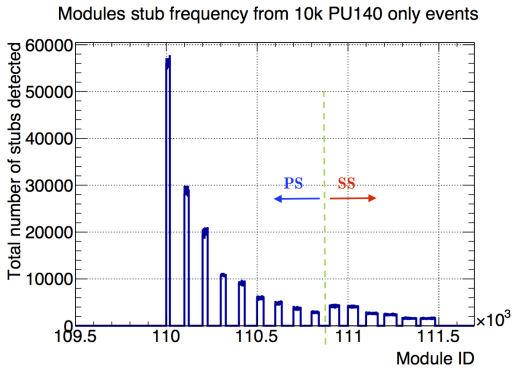

Figure 5.3: Stubs produced in 10k pile up only events with 140 interactions on the 11th layer of the tracker.

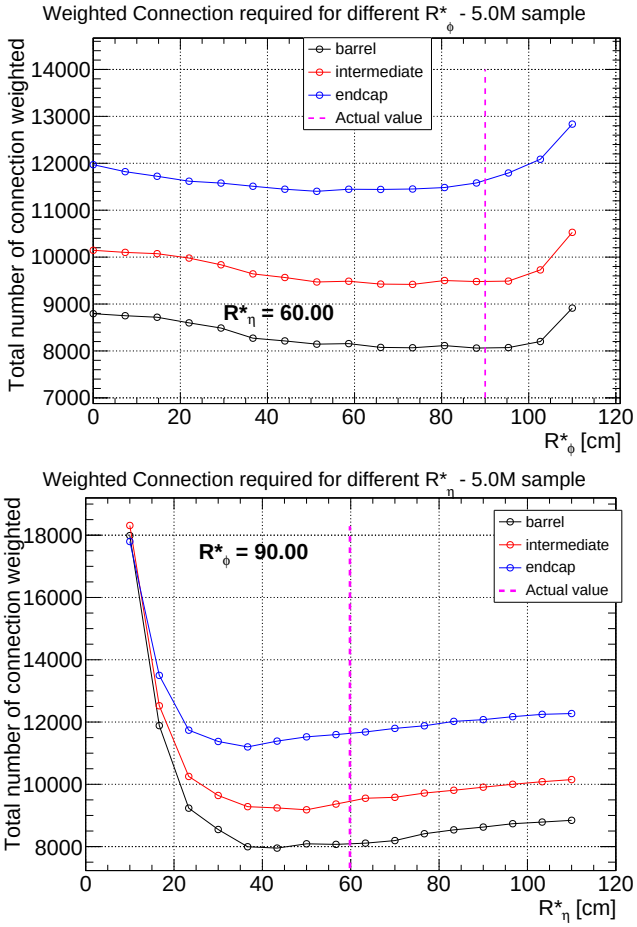

Figure 5.4: R^*_ϕ and R^*_η optimization scan for a TT definition with $|v_z| < 15$ cm and $P_T > 3$ GeV. In while scanning in R^*_ϕ, R^*_η was fixed to 60 cm, whereas in the opposite case the value was 90 cm.
For each of the three zones of the tracker the total weights sum for a the single TT is summed together.

The optimal values are $R^*_\phi \simeq 60$cm and $R^*_\eta \simeq 40$cm which reduce of about 10%, from starting values, the number of stubs to be delivered.

6 Modules - acceptance trade-off

The target of this section is to provide a trial of reducing the number of modules connected to a TT to the detriment of acceptance. The bottom line is to remove from the TT some modules, the ones less involved, with the purpose to find a compromise between acceptance and amount of information to be analyzed.

To purse this target the strategy adopted consists in determine which modules is worth to remove from each TT and evaluate the detection efficiency as a function of the number of modules removed: the optimal point will be the one with the lower number of modules and that match the efficiency request.

Practically speaking, the study has proceeded as follow:

- An inclusive simulation of muons only events has been performed: 4π muons shooting has been used as test to check which modules in each TT are less necessary to pattern recognition.
- For each TT, the number of stubs detected in each module is recorded
- Inside each layer, modules are sorted by the number of stubs detected. Popularity, defined as stubs detected in a module divided by the total number of stubs detected in the belonging layer, is used.
- For each layer, include modules in TT list from the most popular until the point in which the sum of popularity of the modules included reach a given threshold.
- Compute efficiency using, as denominator, all tracks whose parameters fall inside the TT phase-space region and, as numerator, tracks that have at least 5(6) stubs in modules belonging to the TT in 5(6) different layers.

In principle this study can be conducted for each trigger tower but, considering the symmetries, is it possible to analyze only three different cases which corresponds to one tower in each region: Central (TT25), intermediate (TT33) and forward (TT41). The efficiency scan for them is presented in fig. 6.1. First of all it is important to notice that, even with a threshold of 1 (which is all modules included), the efficiency in lower than one: since in simulation is not included the silicon efficiency (at least to the best of my knowledge), this is due to geometry effect for which some muons may not produce any stubs id the pass near the border of 2 modules.

What is more there is a relevant discrepancy between the 5/6 and 6/6 efficiency for the intermediate region and that is due to the fact that tracks belonging to this region, in particular the ones with $0.9 \lesssim \eta \lesssim 1$, do not go through 6 layers because of the layer position which has some acceptance holes (see fig. 1.1).

In any case, what it is easy to understand from this study is that, if it is possible to bear losses of less than 1% in acceptance, than the number of modules that has to be connected to a TT may go down of order 10%.

Effects on connections per modules - The modules connection reduction method introduced before effects also the fan-out distribution because modules less popular are more likely to be near the border and in particular the ones shared by more than one tower.
In fig. 6.2 is shown the effect on the fan-out distribution made by different popularity cuts. From there

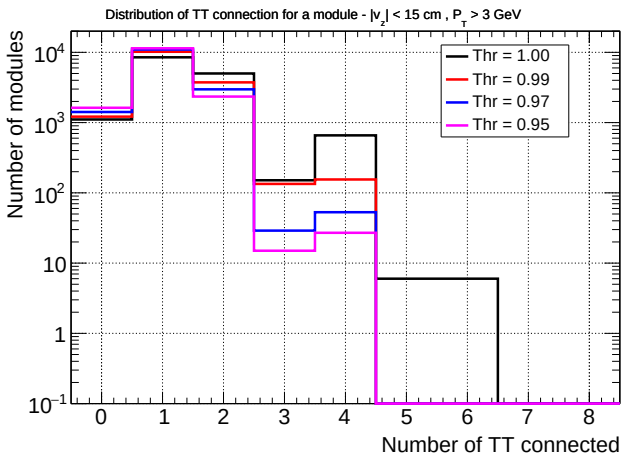

Figure 6.2: Fan-out distribution for different popularity threshold chosen during the TT definition.

it is important to notice that the overall shape of the

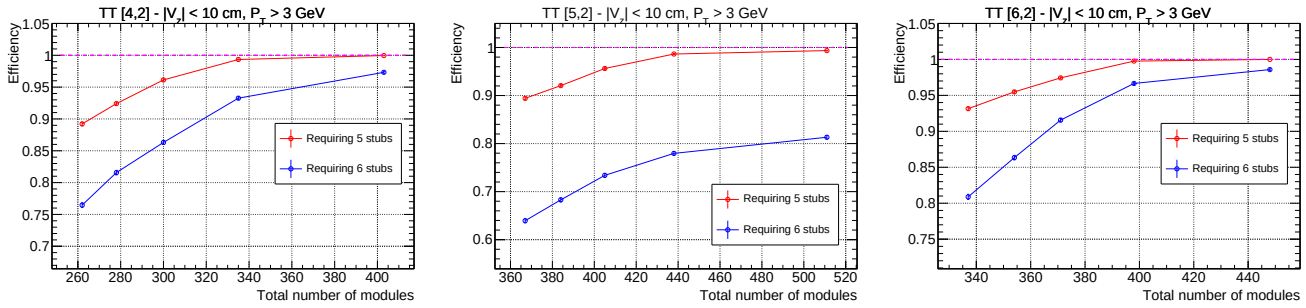

Figure 6.1: Efficiency 6 out of 6 and 5 out of 6 as a function of number of modules in 3 different towers, one for each sector.
Thresholds values used are {1., 0.99, 0.97, 0.95, 0.93} and correspond to the five points in the three graphs.

distribution does not change, especially for low fan-out, but the high fan-out modules (5 and 6), which required the purge procedure explained in sec. 4.1, already disappear for a threshold of 99%.

7 Tracking trigger simulation

In this section the results of the simulation of the tracking process, both AM pattern recognition and track fitting, is presented.
This study has been performed using the new definition of the trigger tower introduced in this document. To perform the definition using the full simulation approach the following parameters has been fixed:

- 48 towers from 6 equal divisions in η^* and 8 in ϕ^*
- $R^*_\phi = 90$ cm
- $R^*_\eta = 60$ cm
- Popularity threshold equal to 1 (that is no popularity cut applied)
- P_T acceptance up to 3 GeV
- v_z acceptance defined by $|v_z| < 15$ cm
- Simulation sample of 20M 4π single muon cleaned events

The definition is, in concrete terms, made of 2 `.csv` file containing the module and the physical boundaries for each TT.

Once defined the TT, AM pattern banks for the new towers are generated using the standard L1TTUpgrades simulation code, properly modified to take in account TT definition modifications.
As state of art, only for TT25 a full simulation and a complete pattern bank is available whereas partial pattern bank (PB) have been generate for TT17 and TT26 in order to study the effect near the border edges in η and ϕ.

The pattern banks generated have been used to run the full simulation of the trigger and to evaluate the performance of the new definition.
Efficiency, average output roads (strictly correlated with latency) and resolution studies are presented what follows.

7.1 Pattern bank

Pattern banks are the core of the AM approach and may vary in size according to the purpose and settings. They depend also on the super strips (SS) configuration and, in this study, six different SS configuration have been taken in account.

The starting point for PB generation is a row sample of single muon events and no pile up. The margin in phase-space of this sample are, on purpose, wider than the ones of the TT25, which therefore fully contained.
This choice is useful, even if reduce the effective size of the sample because not all the particle inside the sample belongs to the TT25 and can be used to generate patterns, because can be used as well to generate pattern banks for neighbor tower near the border with TT25 and because can be reused in future if different configurations are taken in account.

The procedure of getting from the raw sample the proper training sample, called *shrinking*, is performed together with the *stubs cleaning*, which consists in removing redundant stubs. The training

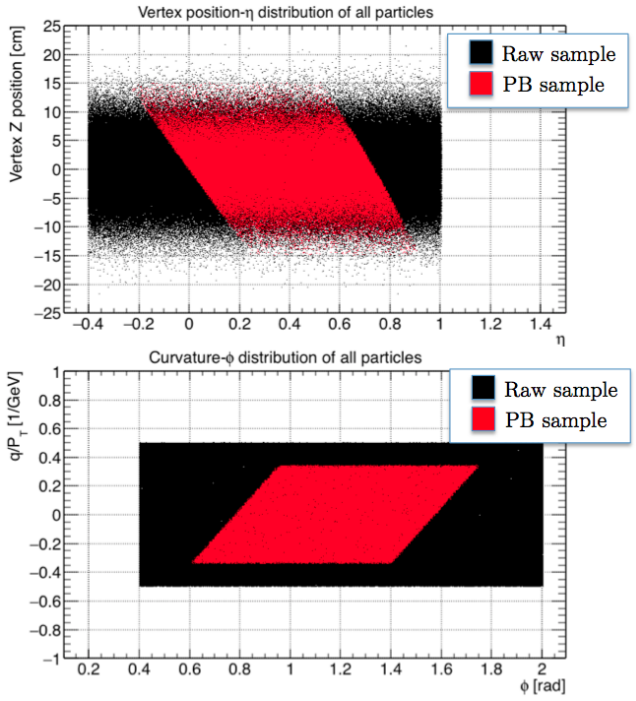

Figure 7.1: Comparison between raw (black) and training sample (red) phase-space for TT25. With this choice of boundaries only about the 20% of the raw sample particles fall inside TT25 definition phase-space and are therefore included in the training sample.

sample so obtained includes only particle that are inside the phase-space assigned to the related tower, that in this case is TT25. In general the shrinking and cleaning procedure of the same raw sample has been applied, with the proper modifications, to obtain als o the partial pattern bank for TT17 and TT26.

To reduce the PB dimensions not all the pattern are stored in the AM but, following a procedure similar to sec. 6 is applied: patterns are sorted by popularity, that is the number of particles inside the training sample that fire the pattern, and than the pattern bank is build using the most popular pattern up to the point in which the sum of the popularity of included patterns is the 95% of the whole training sample.
In fig. 7.2 a table with the banks size for different super strips configuration is reported. In order to give sense to this procedure of pattern bank generation, it is very important that all the included patterns have a decent popularity. This is to reduce both statistical fluctuation and pattern bank size: if the training sample is too small and the less popular included pattern has low popularity (smaller than 3) the previous algorithm tends to include more pattern and the ones included not always are the most relevant ones.
To generates PB for TT25, at the beginning we used a sample of only 68M events but than we increased to 200M to reach satisfing popularity for the *nz8* configuration.

	NEW		OLD	
SSConfig	Bank size 95%	Popularity	Bank size 95%	Diff
sf1_nz4_L5x2	1.40E+06	13	1.31E+06	6.8%
sf1_nz6_L5x2	3.02E+06	6	2.85E+06	5.9%
sf1_nz8_L5x2	5.53E+06	3	4.95E+06	11.6%
sf1_nz4_L5x2_L10x2	1.07E+06	18	1.00E+06	6.9%
sf1_nz6_L5x2_L10x2	2.31E+06	9	2.18E+06	5.8%
sf1_nz8_L5x2_L10x2	4.22E+06	5	3.77E+06	11.7%

Figure 7.2: Pattern bank size for TT25 using different super-strips configurations. Training sample dimension 200M single muons events.

7.2 Number of roads

An important parameter to evaluate trigger performances is the number of roads fired per event. This number is directly correlated with trigger latency once known the average time needed for the system to process each road.

In order to pursue the previous goal the average number of roads in events with pile up 140 and a $t\bar{t}$ couple in final states has been generated and passed through the full simulation of detector and tracking trigger. This kind of event has been chosen because of the facts that these events are physical relevant and among the most crowded.
In fig. 7.3 some statistics are reported. It is clear

TTbar + PU140 events						
	NEW		OLD			
SSConfig	Mean roads	road 95%	Mean roads	Diff	roads 95%	Diff
sf1_nz4_L5x2	28.00	108.1				
sf1_nz6_L5x2	22.85	89.08	22	3.9%	90.8	-1.9%
sf1_nz8_L5x2	21.19	85.31	20.6	2.9%	82.8	3.0%
sf1_nz4_L5x2_L10x2	29.40	115.3				
sf1_nz6_L5x2_L10x2	23.75	94.07	22.8	4.2%	93.2	0.9%
sf1_nz8_L5x2_L10x2	21.96	91.86	21.1	4.1%	85.3	7.7%

Figure 7.3: Comparison between the number of roads produced using old and the new definition. Old data from [5].

that with the new TT definition the number of roads produce slightly increase but it essentially remains of the same order ($\sim$ 20 per event). In the same table is reported the "roads 95%" which is the value for which the 95% of events produce a

smaller number of roads: this is important because gives an upper bound on the total roads that needs to be processed.

For what is above an interesting figure of merit to compare different configurations is the number of roads versus PB size (fig. 7.4). In these kinds

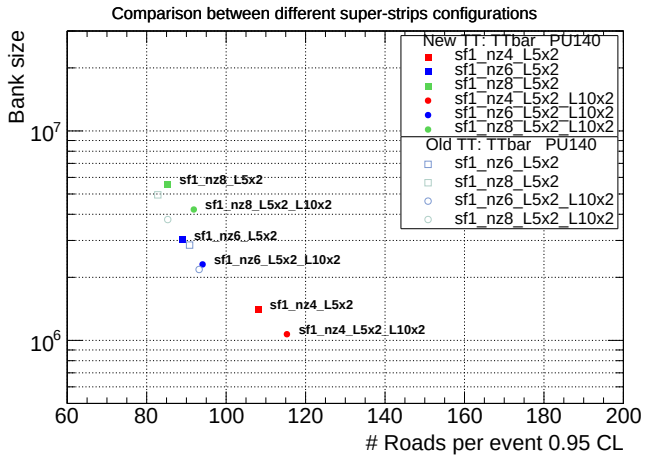

Figure 7.4: Bank size versus number of roads from a sample of 2000 $t\bar{t}$ events for old and new TT definition and different super strips configuration. Old data from [5].

of plots the workload between AM and FPGA is explored and for this the best case scenario is the one closer to the origin.

7.3 Efficiency

The second step in performance evaluation is the efficiency and resolution study. In the meanwhile this study has been used to verify the effective TT dimension and turn-ons near the phase-space edges. The sample used for the purpose of this section is the made of 100k events of single muon "test" (as single muon but withdelta rays enabled at MC level) with no pile up.

Unless different specified the efficiency is meant to be defined:

- For TT global efficiency (synthetic, 5oo6): the denominator is the number of MC tracks with parameters inside the TT definition phase-space, and the numerator, is the number of MC tracks in denominator that have a matching in fitted tracks, reconstructed from roads with at least 5 stubs out of 6 belonging to a pattern.
- For efficiency scan in a given variable: The analysis is performed binned in values of the given variable (also called blind). The denominator is the number of MC tracks with all the parameters, except the blinded one, inside the TT definition phase-space, and the numerator, is the number of MC tracks in denominator that have a matching in fitted tracks, reconstructed from roads with at least 5 stubs out of 6 belonging to a pattern.

For example, for the efficiency scan in P_T, to be in the denominator tracks must be inside TT phase-space in ϕ^*, η^* and v_z, but no requirements on P_T is applied.

Efficiency study in P_T and ϕ^* are dependent on the super strips configuration only if the size in ϕ is changed but, since we used always `sf1_L5X2` for this study, that is not the case.
In fig. 7.5 and fig. 7.6 the efficnecy as a function of P_T and ϕ^* is presented. Relevant features of

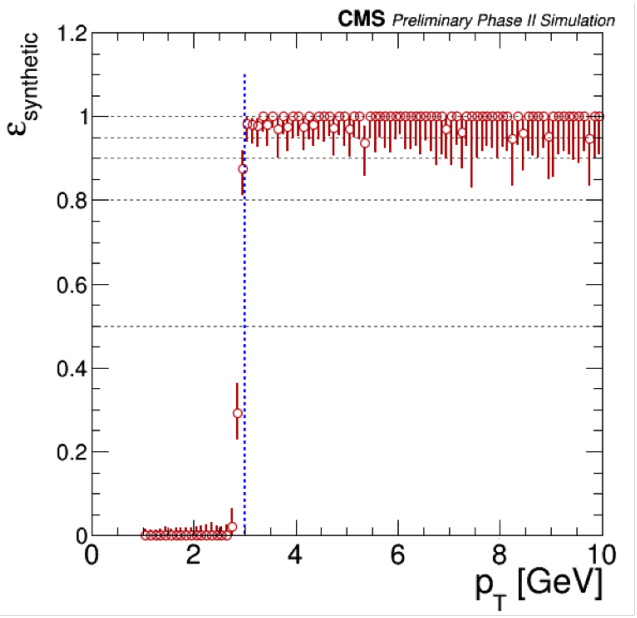

Figure 7.5: Efficiency as a function of P_T for sf1_nz4_L5x2. TT definition phase space border are marked with blue dashed lines.

this new definition is the very sharp turn on, the almost null efficiency outside the TT and the very high efficiency inside the TT ($99.2 \pm 0.1\%$).

About v_z and η^* the differences between the various configuration starts to be relevant.
Efficiency as function of tracks v_z (fig. A.4) is affected by low statistic near turn-on boundaries be-

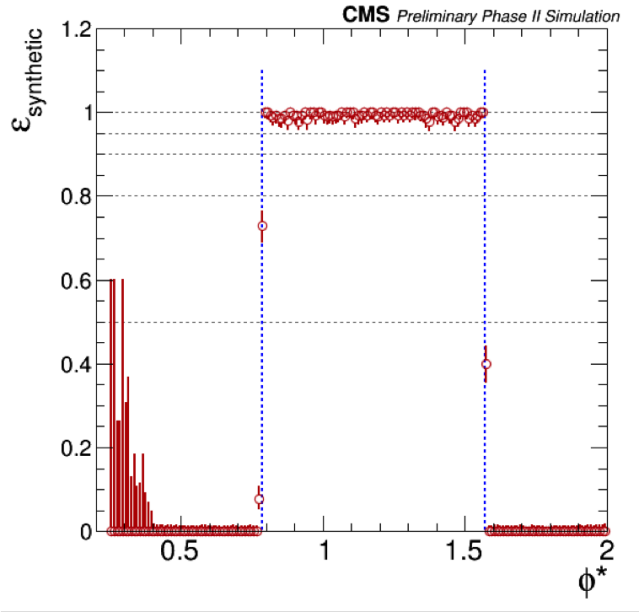

Figure 7.6: Efficiency as a function of ϕ^* for sf1_nz4_L5x2. TT definition phase space border are marked with blue dashed lines.

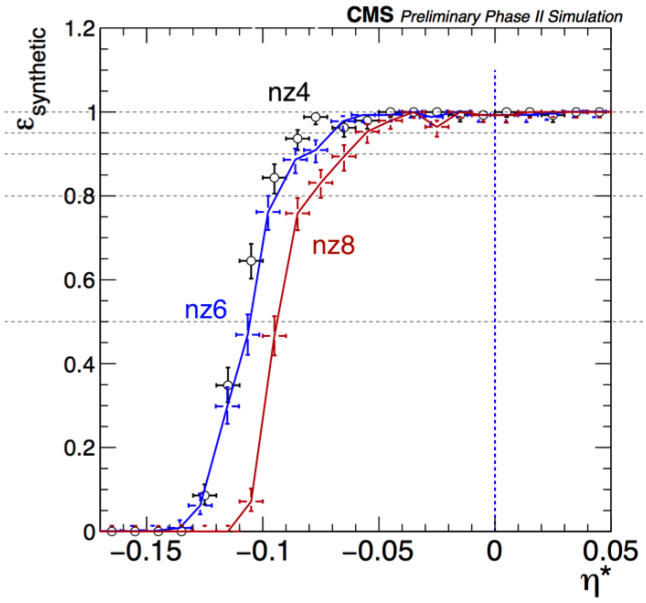

Figure 7.7: Comparison between the efficiency as a function of η^* evaluated with different super strips configurations.

cause the TT v_Z border is wide (more than 3σ out). Anyway turn on is smoothed by large z division in super strips but the average efficiency inside TT is still $99.2 \pm 0.1\%$, because of course it does not depend on the order of the requirements. Peculiar characteristic are some efficiency drops given by modules geometry.

The most interesting case is the efficiency scan as function of η^* (fig. 7.7). In this case the finite size of the super strip division in the z coordinate is relevant and cause an early turn on and effective efficiency wider than definition.
What is more this is supported by the fact that reducing the dimension of the z division, i.e.increasing nz, the over coverage decrease.
The bottom line of thie effect is that particle outside TT definition may match roads: indeed super strips are defined taking the maximum range of interesting particles on TT, and that is why a single training particle does not lay on the border of every layer, but outside TT particles can do that. This is why pattern bank acceptance is actually broader than training sample phase space and this effect is lager with increasing super strips size. It is possible to verify that this happens in ϕ too but it is far smaller because the number of division is much larger in that dimension.

It is important to notice that, even if this is an intrinsic way to generate duplicate fitted tracks, it has a lower impact on 6 out of 6 efficiency (fig. 7.8). In 6oo6 efficiency indeed the turn on is much more

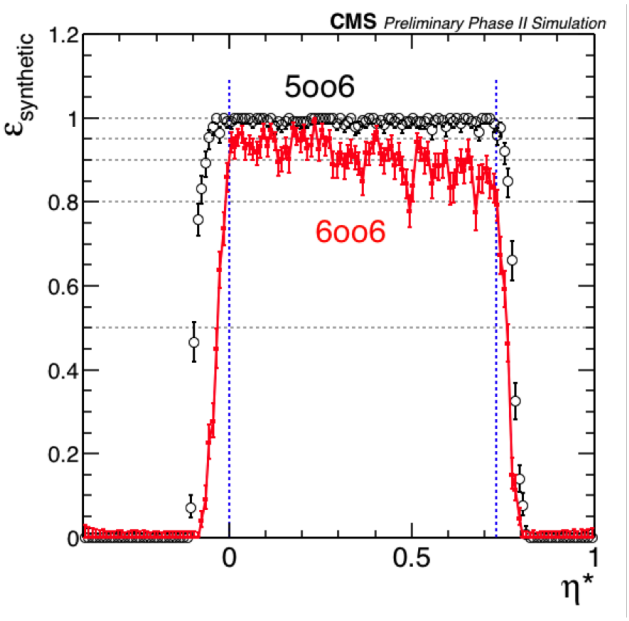

Figure 7.8: Comparison between the efficiency 6oo6 and 5oo6 as a function of η^* evaluated for nz4.

sharper and closer to the TT definition boundaries. It is of course true that 6oo6 efficiency has a lower average efficiency (inside TT $90.8 \pm 0.3\%$) because

of the PB 95% cut (Si efficiency is set practically to 1 in MC) and of module geometry effect which cause efficiency drops and the slope towards high η.

Finally also the behavior of efficiency near the has been studied.
To do that the same raw sample used for TT25 has been used to generate partial training sample for neighbor tower. This procedure, possible thanks to the choice of using a raw sample larger than the strict TT definition, lead, in general, to incomplete PB for the neighbor tower but that are in any case complete near the border region with TT25.
From this study (graphs in fig. A.7 and fig. A.6) it is clear that the new definition does not suffer from trigger tower border issues and instead rise the problem of over efficiency in the η^* border.

7.4 Resolution

As last step of performance evaluation the tracking trigger resolution as been evaluated. Results are shown in fig. 7.9.

The output resolution in very good, about 1%, and the efficiency scan in P_T and η are in full agreement with the expectation.

8 Conclusions

A new definition of trigger tower for the phase II CMS tracking trigger has been developed, together with a dedicated code to generate definition files and perform definition parameters optimization.

The optimization studies for R^* brought to the choice of the value of $R^*_\eta = 60$ cm and $R^*_\phi = 90$ cm. Further studies demonstrate that a v_Z acceptance of 15 cm is worth and can operate with a number of modules per trigger tower between 400 and 500. Efficiency studies shown that the mismatching present in tkLayout is completely resolved and nor low efficiency inside TT neither edges problems are present. Finally, the P_T resolution has been confirmed to be order of 1 per cent.

9 Acknowledgments

I wish to thank my supervisor Luciano Ristori for many useful discussions and guidelines and Jia Fu Low and Sergo Jindariani for the support and help.

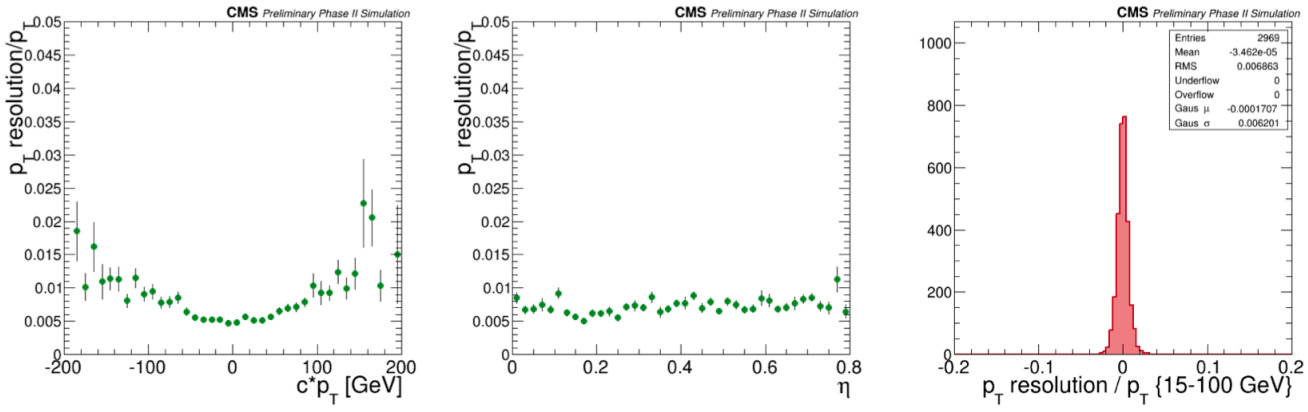

Figure 7.9: Resolution on P_T results.

A Appendix - Further Graphics

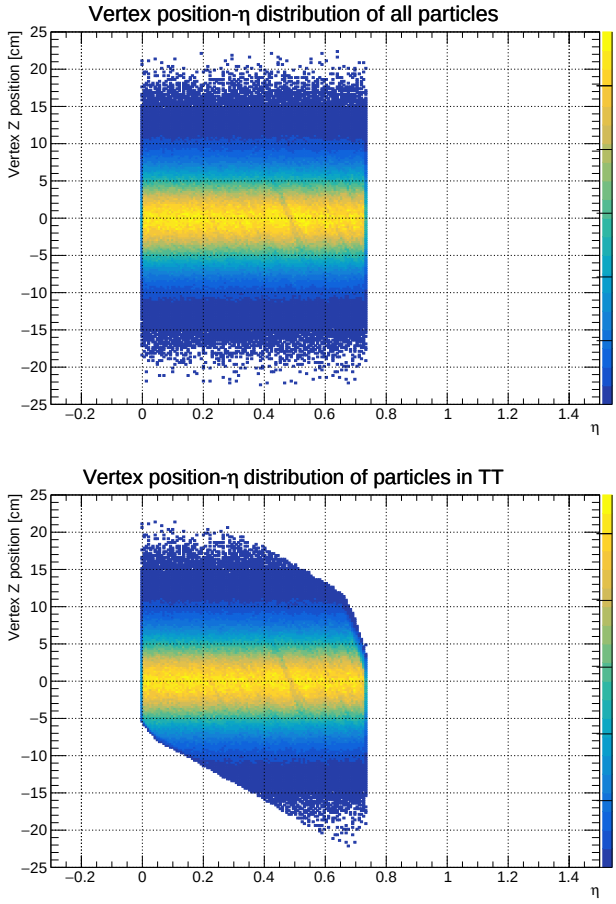

Figure A.1: Comparison between the v_z, η phase space of TT27 training sample and definition.
Top: All tracks inside training sample.
Bottom: All tracks in training sample with at least 6 stubs inside the TT definition.

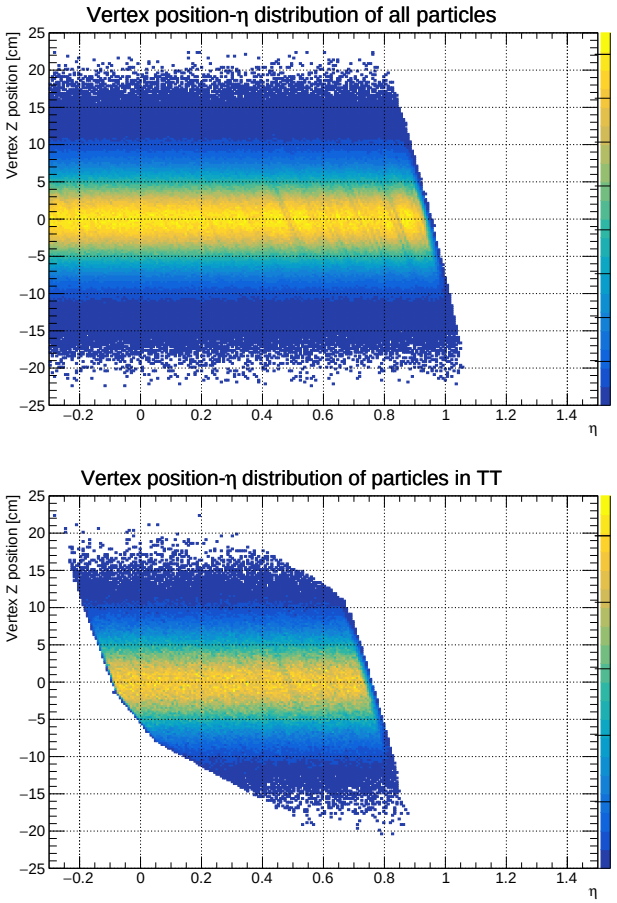

Figure A.2: Comparison between the v_z, η phase space of the global sample and TT27 definition.
Top: All tracks inside global sample.
Bottom: All tracks in global sample with at least 6 stubs inside the TT definition.

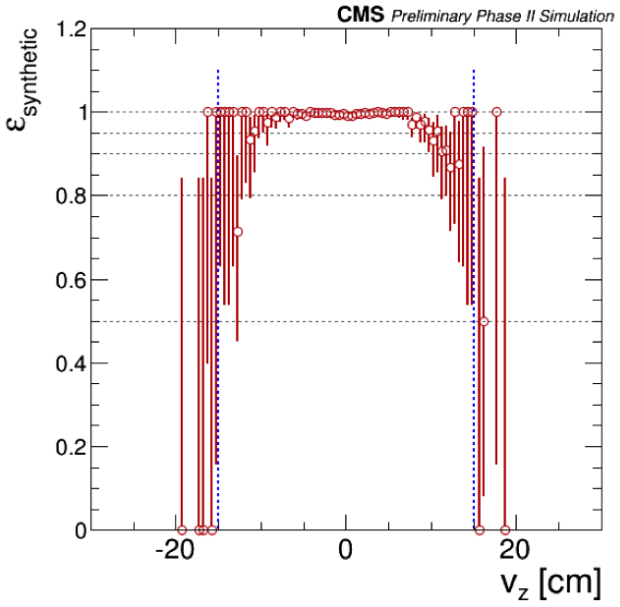

Figure A.4: Efficiency as a function of v_z for the sf1_nz8_L5x2 super strips configuration.

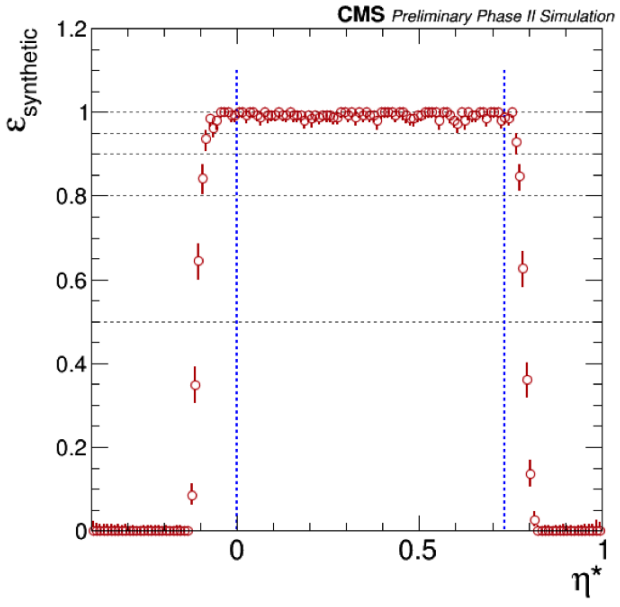

Figure A.5: Efficiency as a function of η^* for the sf1_nz4_L5x2 super strips configuration.

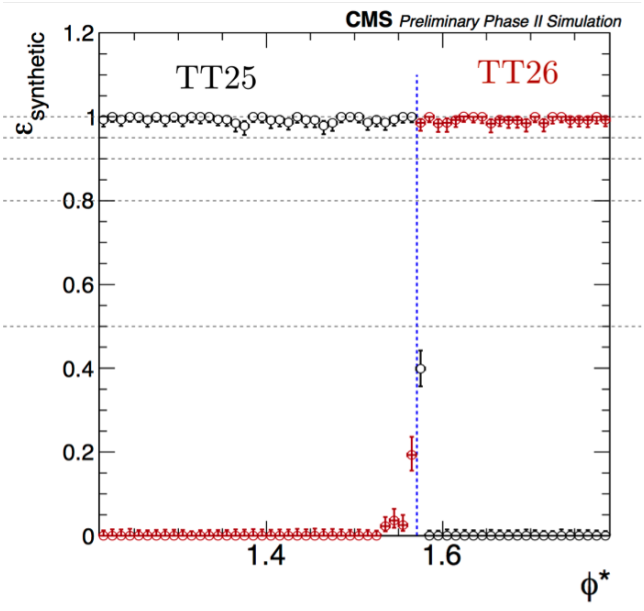

Figure A.6

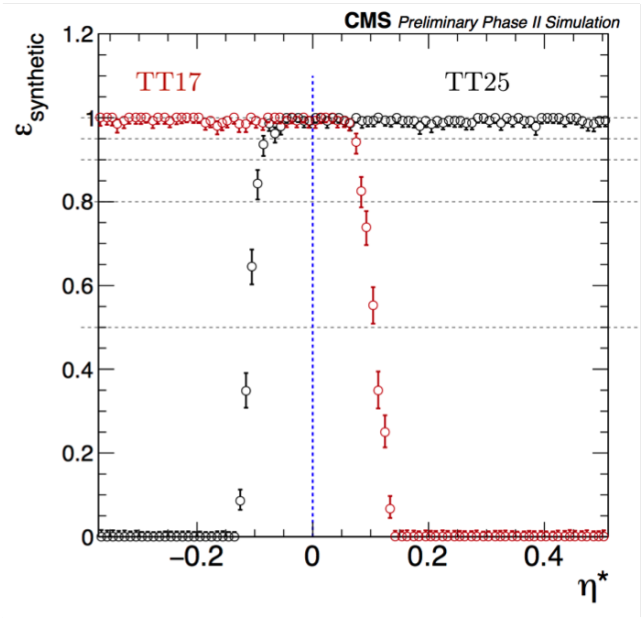

Figure A.7

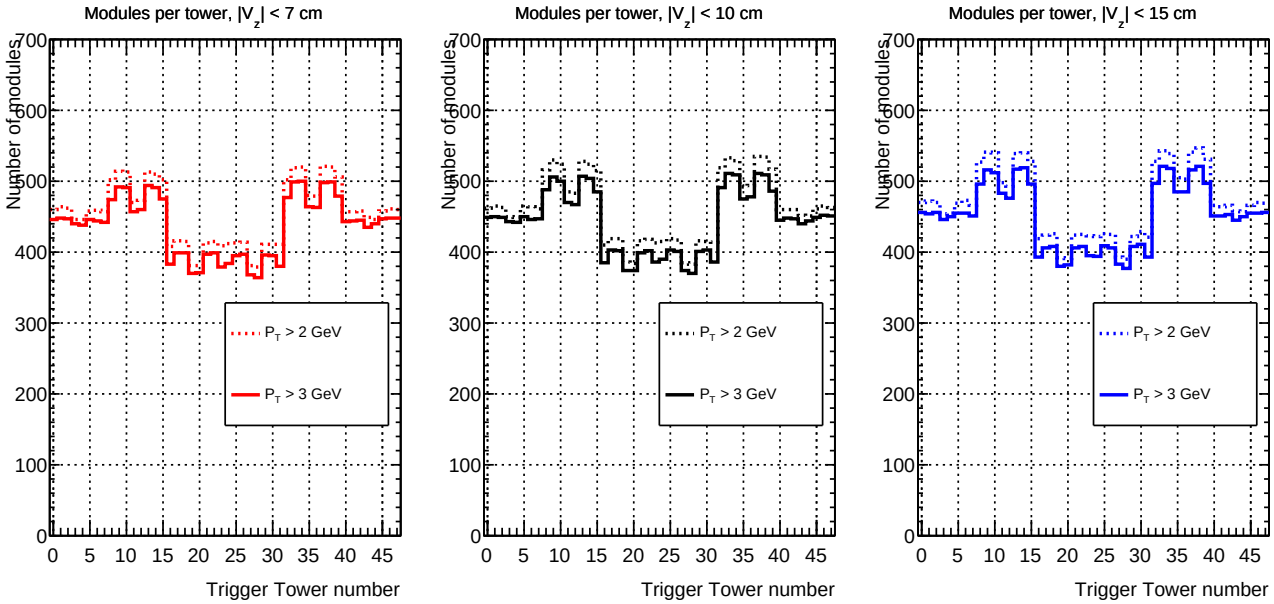

Figure A.3: Number of modules connected to a single TT in various configurations.

B Appendix - Particle propagation

B.1 Transverse plane

Particles propagating in a magnetic field obey the following relations:

$$P_T = cB\rho \Longrightarrow P_T\,[\text{GeV}] \sim 0.3B\rho\,[\text{Tm}]$$

where ρ is the curvature radius. That relation can be re arranged in the form $\rho = \alpha P_T$, with

$$\alpha = \frac{\text{GeV}^{-1}\text{m}}{0.3 \cdot B\,[\text{T}]}$$

Defining ϕ_0 as the angle between the x direction and the particle velocity when it pass at radius $r = sqrtx^2 + y^2$ equal to zero, can be shown that $\phi^*(R^*)$, defined as the angle between the particle position vector in $x - y$ plane at $r = R^*$ and the x direction, obey the following equation:

$$\frac{q}{P_T} = \frac{2\alpha}{R^*}\sin\left(\phi_0 - \phi^*\right)$$

where q is th charge of the particle.

It is interesting also to notice that (fig. B.1), given the curvature radius ρ, for a particle which start with $\phi_0 = 0$ the following holds:

$$r = r_0 + 2\rho\left|\sin\frac{\omega t}{2}\right|$$

where ω is the pulsation and $r_0 = r(t = 0)$. From this, assuming constant velocity along z axis,

$$\phi(z) = \frac{z\text{tg}\theta_0}{\alpha P_T}$$

and

$$\phi(r) = \phi_0 + \text{asin}\,\frac{r}{2\rho}$$

at first order approximation and for small times.

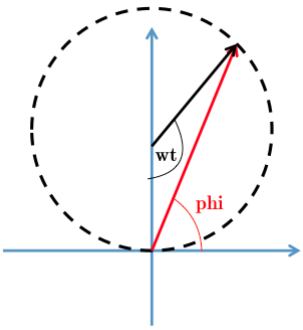

Figure B.1: $x-y$ projection of the trajectory of a charged particle.

B.2 r-z plane

Remembering the definition of pseudo-rapidity

$$\eta = -\ln\left(\text{tg}\,\frac{\theta}{2}\right) = \frac{1}{2}\ln\left(\frac{p + p_L}{p - p_L}\right)$$

where θ is the polar angle between particle position and z axis, is important to keep in mind that

$$\sinh\eta = \text{cotg}\;\theta$$

.

And assuming helicoidal trajectory for the particle in the 3 dimensional space, can be shown that:

$$r(z) = 2\rho\sin\left(\frac{z\text{tg}\theta_0}{2\rho}\right)$$

$$z(r) = z_0 + \frac{2\rho}{\text{tg}\;\theta_0}\text{asin}\left(\frac{r}{2\rho}\right)$$

at all orders. From the last two, and remembering $\text{tg}\theta(r) = \dfrac{r}{z(r)}$, can be derived that

$$\eta(r) = \text{asin}\left[\frac{z_0 + \frac{2\rho}{\text{tg}\theta_0}\text{asin}\left(\frac{r}{2\rho}\right)}{r}\right]$$

References

[1] K. A. Olive *et al.* [Particle Data Group Collaboration], Chin. Phys. C **38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001

[2] Sabes, David. "L1 track triggering with associative memory for the CMS HL-LHC tracker." Journal of Instrumentation 9.11 (2014): C11014.

[3] F. Morsani et al., The AMchip: A VLSI associative memory for track finding, Nucl. Instrum. Meth. A 315 (1992) 446.

[4] A. Annovi et al., Associative Memory design for the fast track processor (FTK) at ATLAS, IEEE Nucl. Sci. Symp. Conf. Rec. (2011) 141.

[5] "Status of Pattern Bank Optimization" (2016-05-20 Update), Ristori et al.