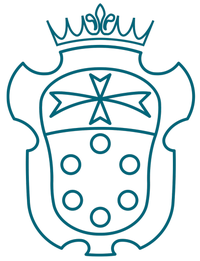
Signal Efficiency and Smearing Studies using FairMUonE Monte Carlo Reconstruction

Pisa MUonE Group Analysis Meeting

Lorenzo Punzi

22/09/2025



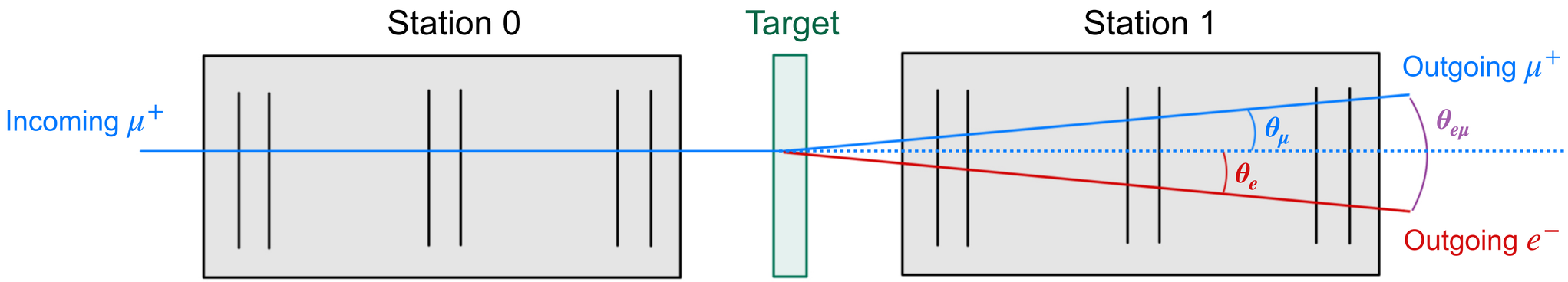
SCUOLA
NORMALE
SUPERIORE

Overview

- The purpose of this work is to study the reconstruction efficiencies of the MUonE tracking detector **using the FairMUonE v14 Monte Carlo simulation and reconstruction** (unless otherwise specified)
- I have only used ***signal*** events so far, forcing elastic scattering in the target
- **`MESMER`** is used as the signal generator at **NNLO** (unless otherwise specified)
- Although the **`MESMER`** generation is weighted, I'm not considering the weights of the events. The weights are only relevant for the cross section, not the efficiency, which is a conditional probability!

Generation Parameters

- Detector configuration provided by Riccardo, consisting in two stations with a 3cm target in between: **`TR2023_geometry_3cmTarget_noECAL_zPosFromMetrology`**



- Beam generator : **`ParticleGunBeamProfile`** (**`beamProfile_narrow`**)
- **`MESMER`** generation is **`weighted`** at NNLO (**`alpha2`**) and **`Eemin`**= 1 GeV
- Alignment file is: **`alignParams_Simulation_MetrologyInput_it10.yaml`**
- **`maxNumberOfSharedHits`** varies between 0, 1, and 2
- **`maxOutlierChi2`** is set to -1, so no cut on the χ^2 is performed

The Different Types of Efficiency

- I have measured the efficiencies of each of the two outgoing tracks and of the whole elastic event in two different ways:
  - ***Global Efficiency*** : The global calculation determines the efficiencies solely conditioned to the presence of all three generated tracks in the Monte Carlo
  - ***Reconstructable Efficiency*** : The calculation of the ***reconstructable*** efficiency is conditioned to an event being "reconstructable" in the first place, per Eugenia's definition ***here***

$$reconstructible$$

Reconstruability
$\geq$ **5** (2X, 2Y, $\geq$1 stereo) GEANT4 hits per each track (incoming μ and outgoing μ, e) + μ_{in} track reconstructed with 6 stubs + $\frac{\chi^2}{ndof} < 2$

- All results have been obtained using **5 million Monte Carlo events generated with `MESMER` and reconstructed with FairMUonE**

Disclaimer!

- For the moment I have mainly considered the tracks from the **`ReconstructedTracks`** container, not **`ReconstructedVertices`**, so ***pre*-vertexing**!
- PID is currently based on the **`track.linkedTrackID`** information
- A track is considered “reconstructed” if shares at least a certain fraction of hits shared with its linked track $\left(\boldsymbol{QLT}\right)$. By default I’ve been using $QLT = 67\,\%$ following Eugenia’s example
- In case of multiple tracks being linked to the same track, the one with the highest fraction of hits shared with the linked track is taken as the reconstructed track at the moment
- All these operations require the knowledge of the Monte Carlo truth, so it cannot be considered a realistic scenario of our actual data reconstruction!

Global Efficiency ε^{glb} Studies

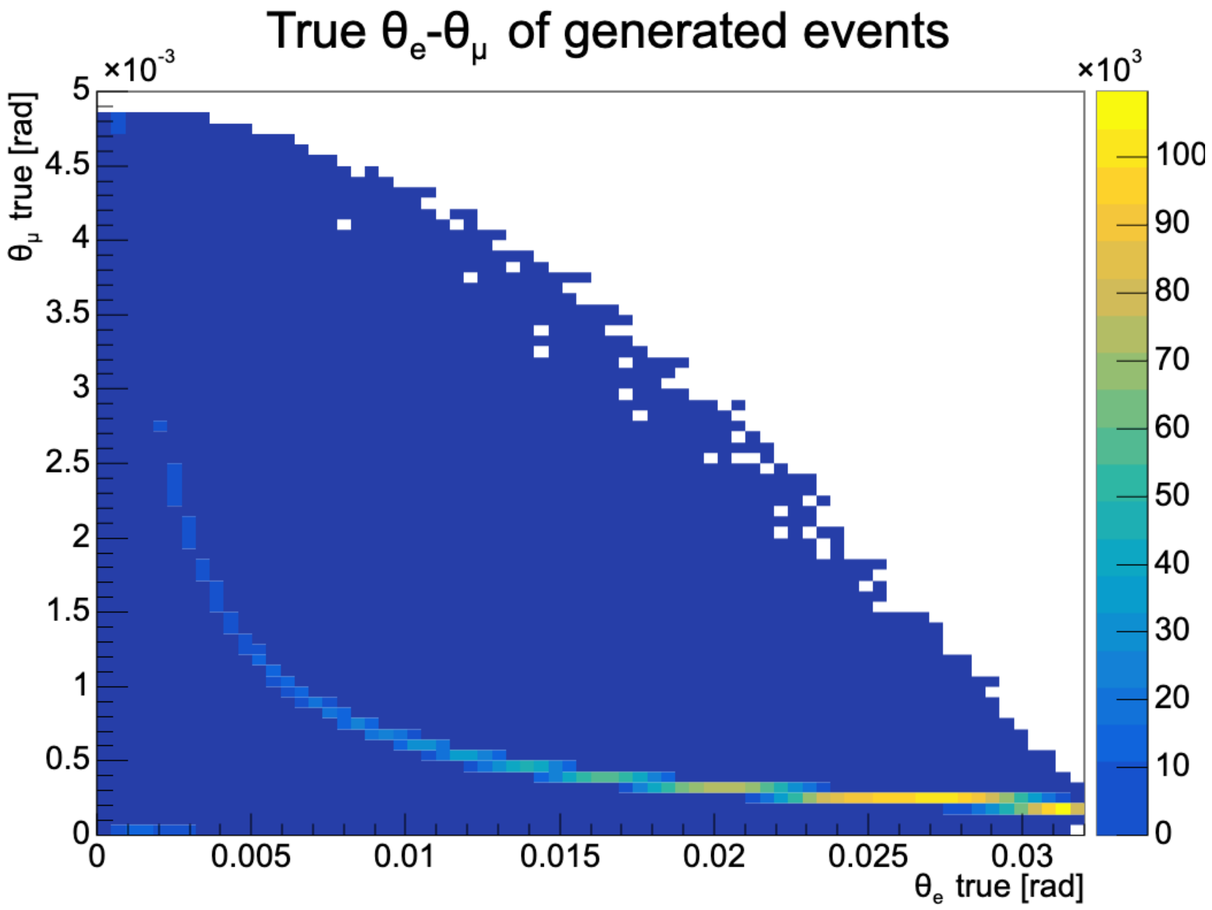
Definition of Global Efficiency ε^{glb}

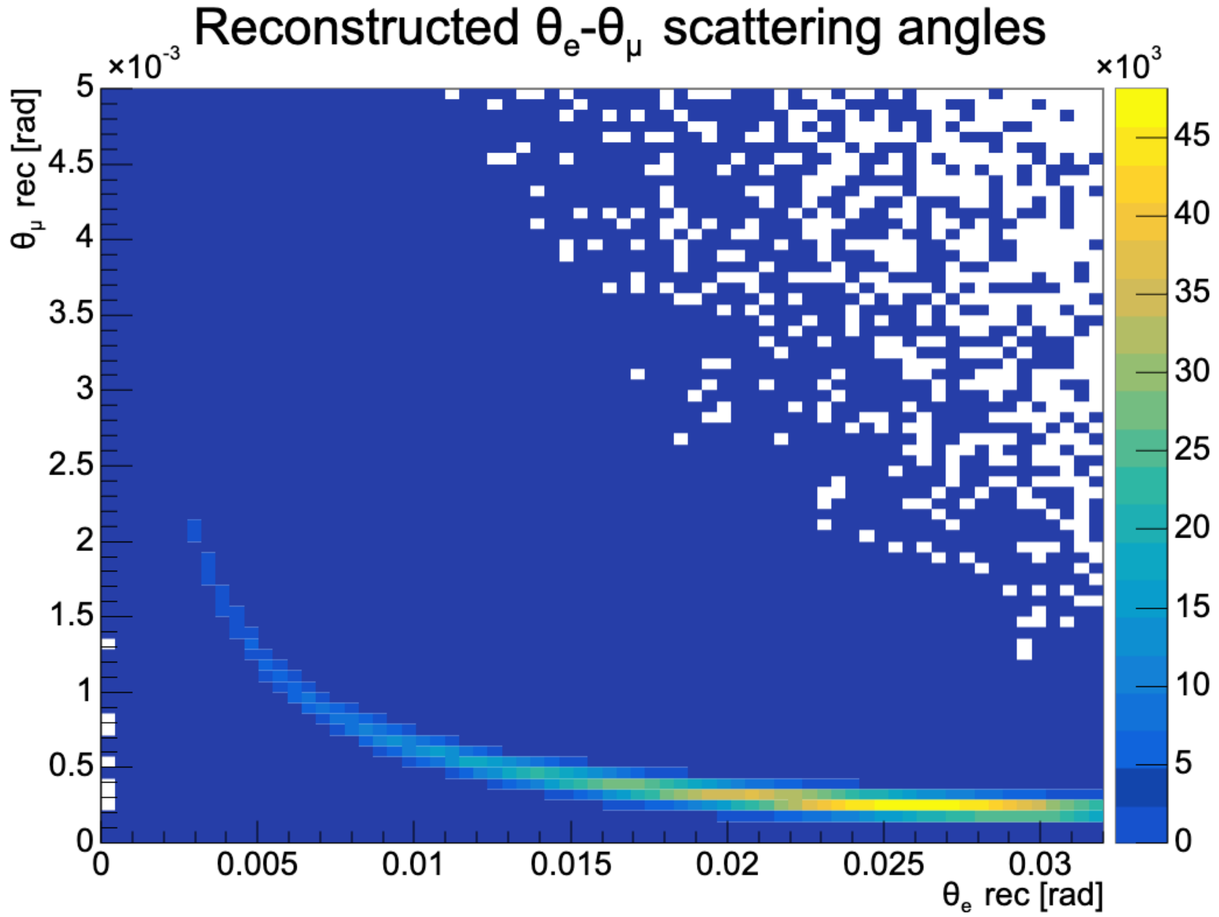
- The global elastic event (ev) efficiency and the two global outgoing track (trk) efficiencies are determined in the following way:

$$\varepsilon_{ev}^{glb} = \frac{N\left[GEN \cap ALLREC\right]}{N\left[GEN\right]} \qquad \varepsilon_{trk}^{glb} = \frac{N\left[GEN \cap TRKREC\right]}{N\left[GEN\right]} \geq \varepsilon_{ev}^{glb}$$

- GEN : Incoming simulated μ_{in}^{+} hits target and both outgoing μ_{out}^{+} and e_{out}^{-} are generated in Monte Carlo
- $ALLREC$: All three tracks of the elastic event are reconstructed
- $TRKREC$: At least incoming μ_{in}^{+} and the track under test are both reconstructed

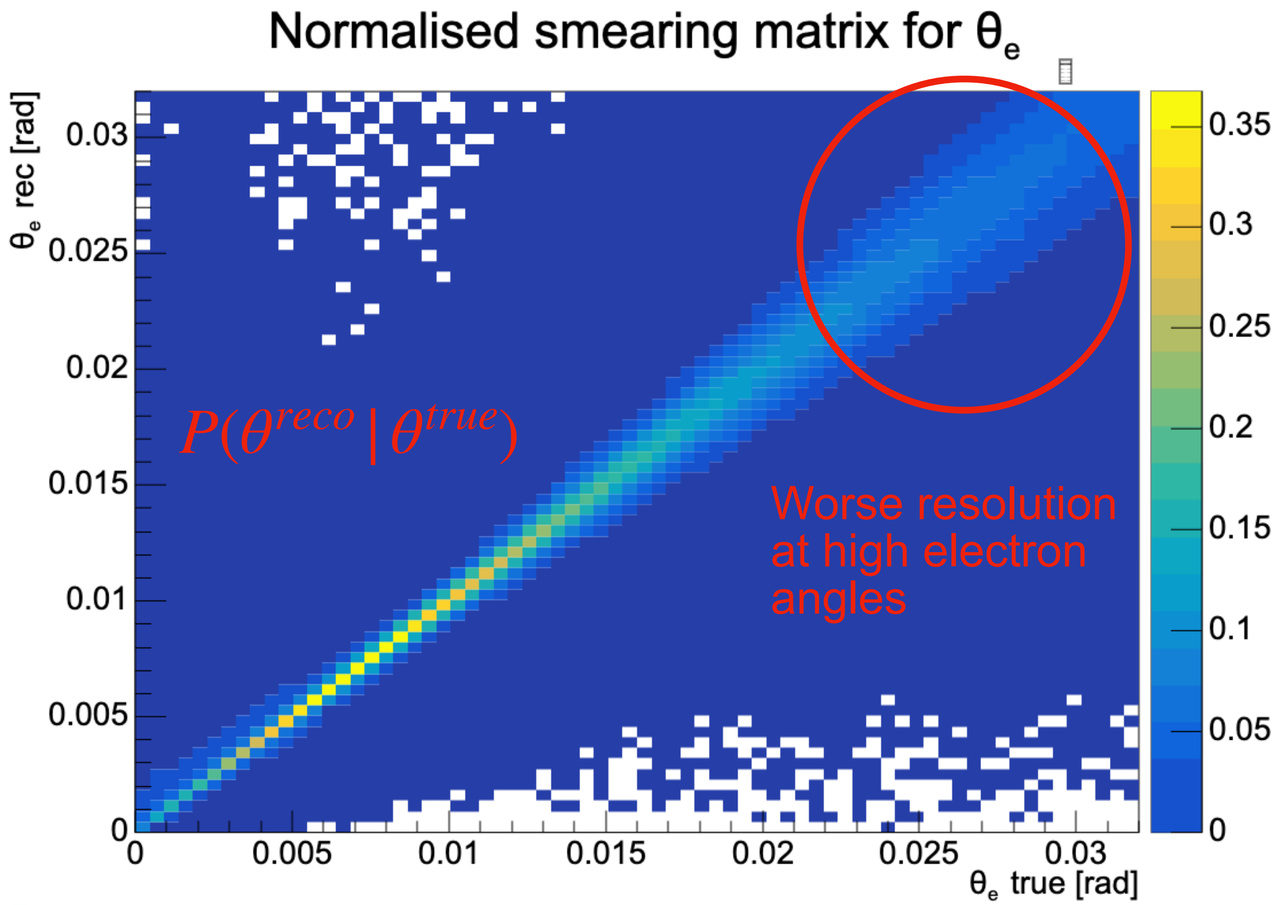
Generated and Reconstructed (θ_e, θ_μ) Distributions

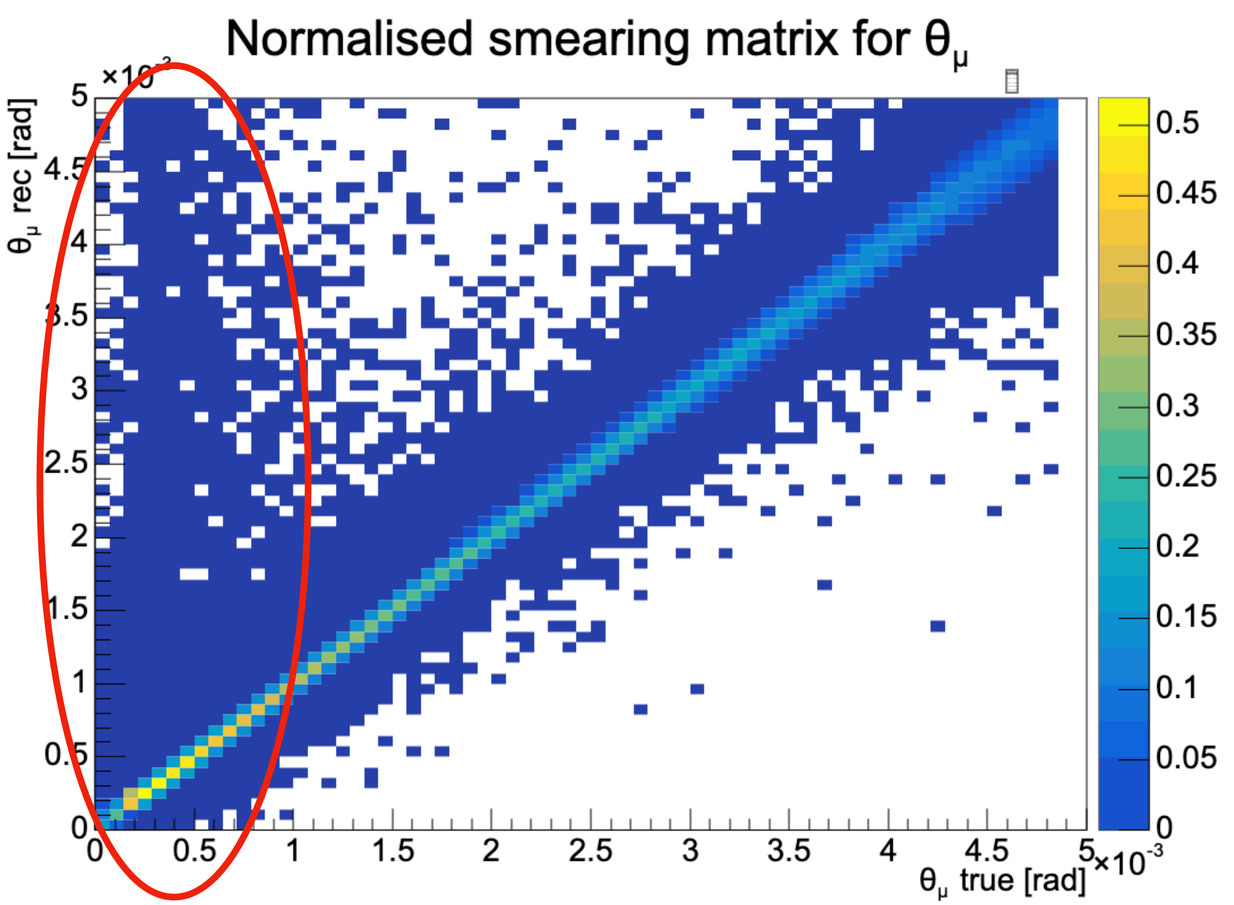




Plots generated using $QLT = 67\,\%$ and **`maxNumberOfSharedHits`** = 0

Normalised Smearing Matrices of θ_e and θ_μ



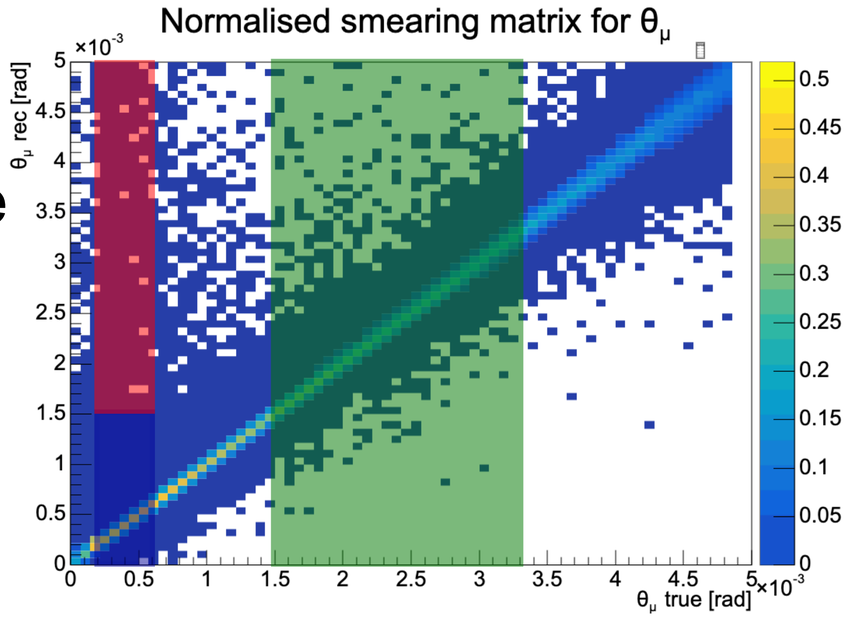


What is the reason for this vertical strip at low θ_μ^{true} ?

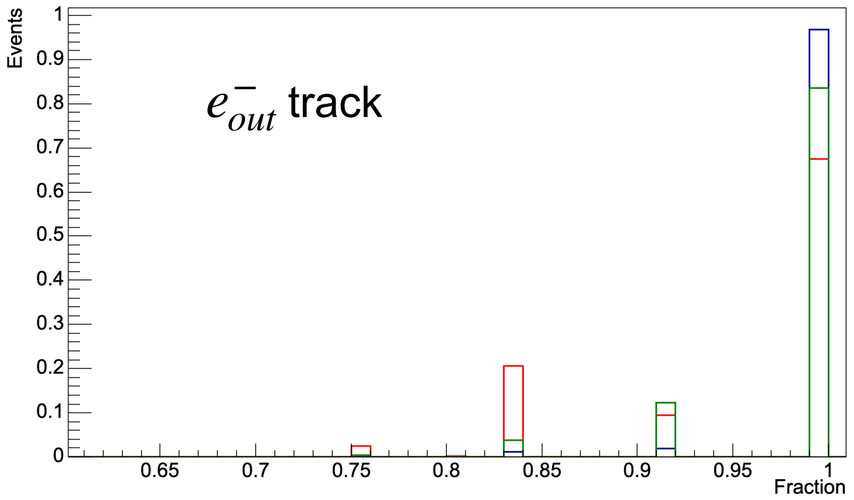
Plots generated using $QLT = 67\,\%$ and **`maxNumberOfSharedHits`** = 0

What is the Vertical Strip?

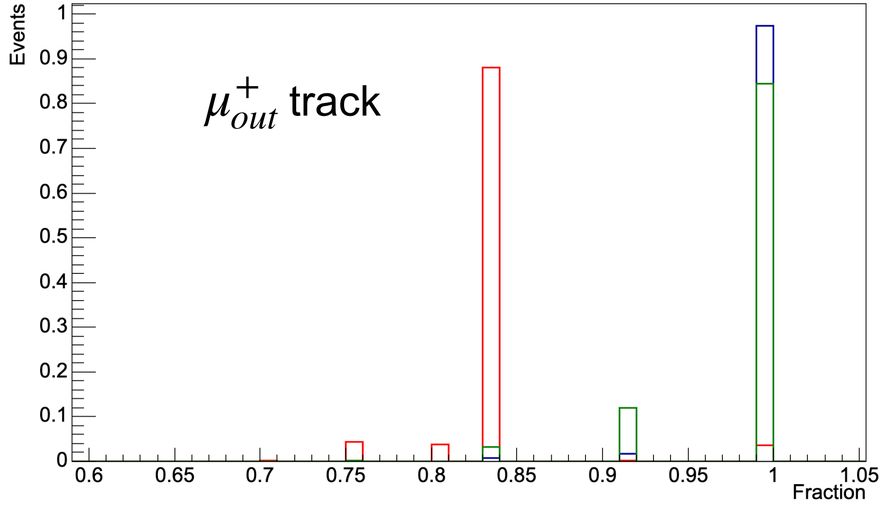
- It looks like the events in the higher part of the strip have much worse fraction of shared hits in common with the linked track with respect to the other two regions



Plots generated using $QLT = 67\ \%$ and **`maxNumberOfSharedHits`** = 0



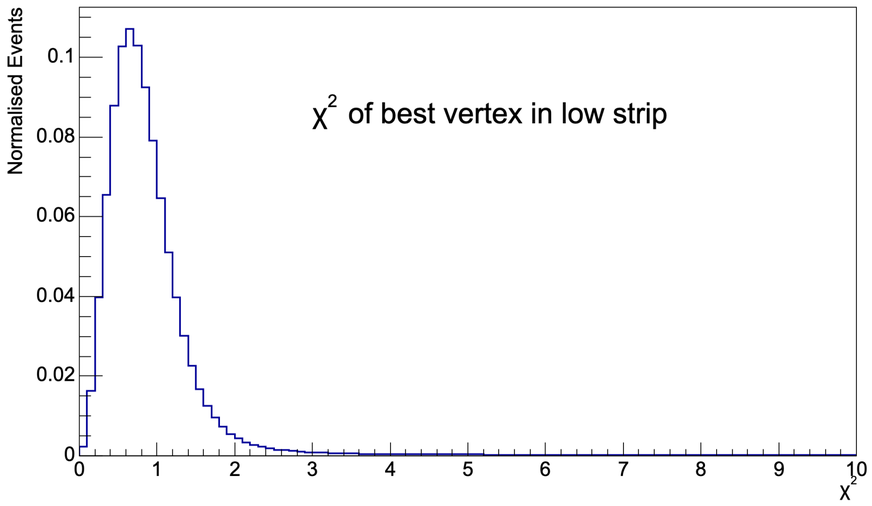
Fraction of hits shared with linked track



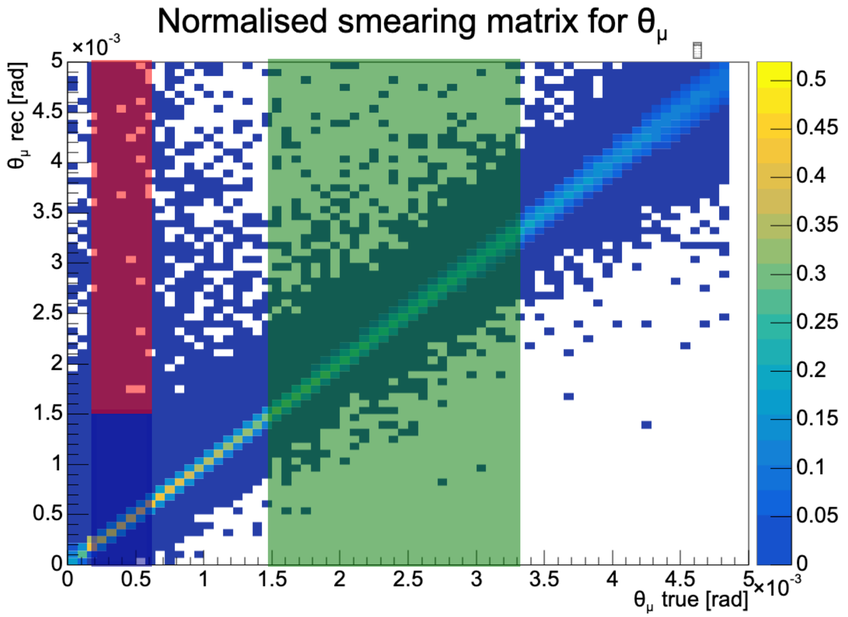
What is the Vertical Strip?

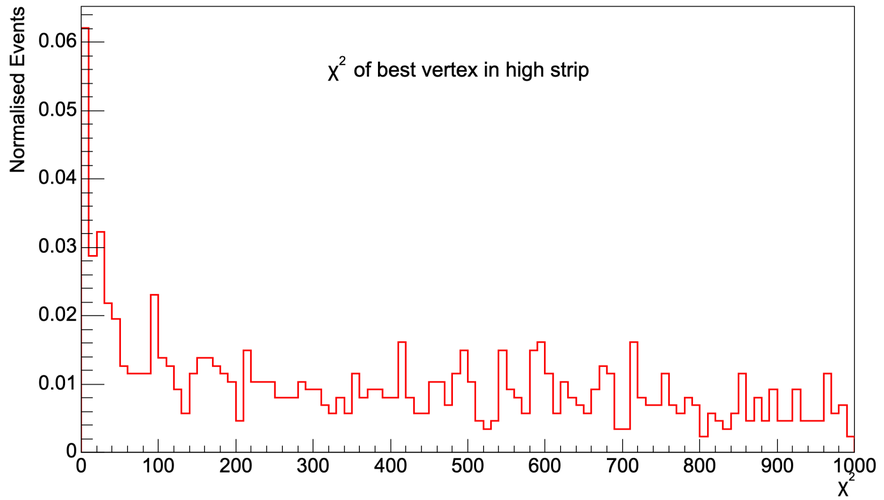
- There is also clearly a much worse $\chi^2/ndof$ of the best vertex for the events in the high part of the strip
- I am currently not performing any cut on the $\chi^2/ndof$ of the vertex, following Eugenia's example. From what I understand this was due to how Multiple Scattering was handled in older versions of FairMUonE

Plots generated using $QLT = 67\ \%$ and **`maxNumberOfSharedHits`** = 0



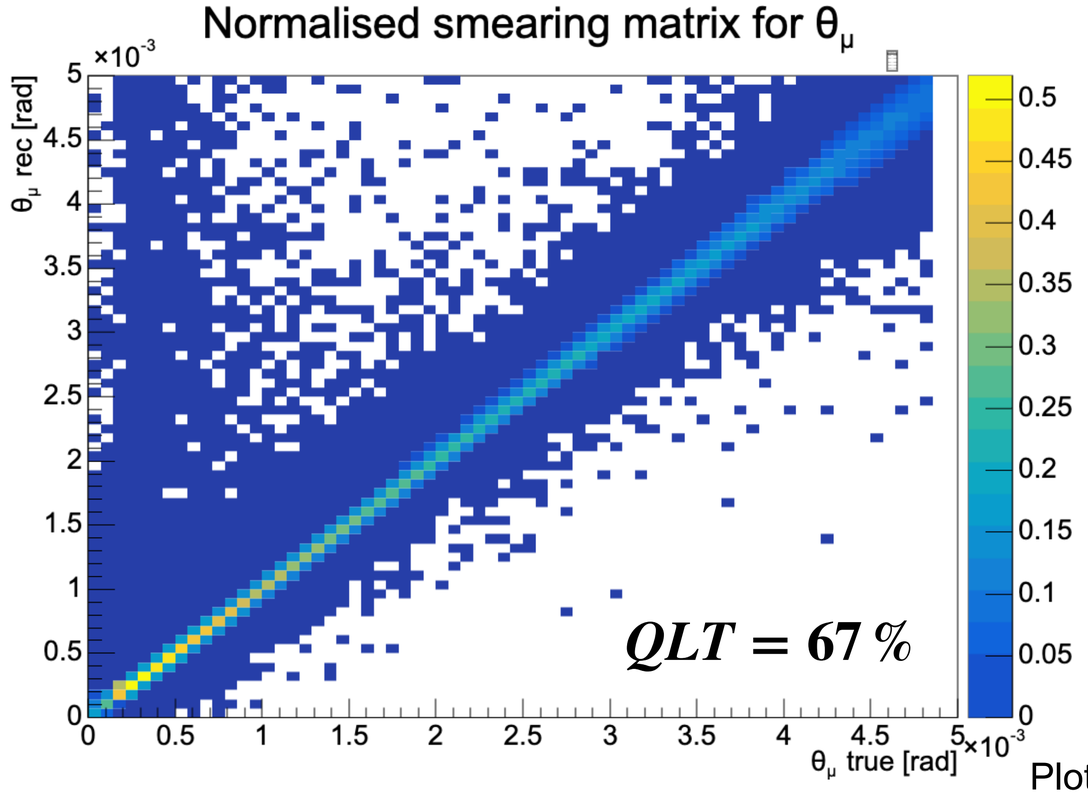
$\chi^2/ndof$ of best vertex distributions

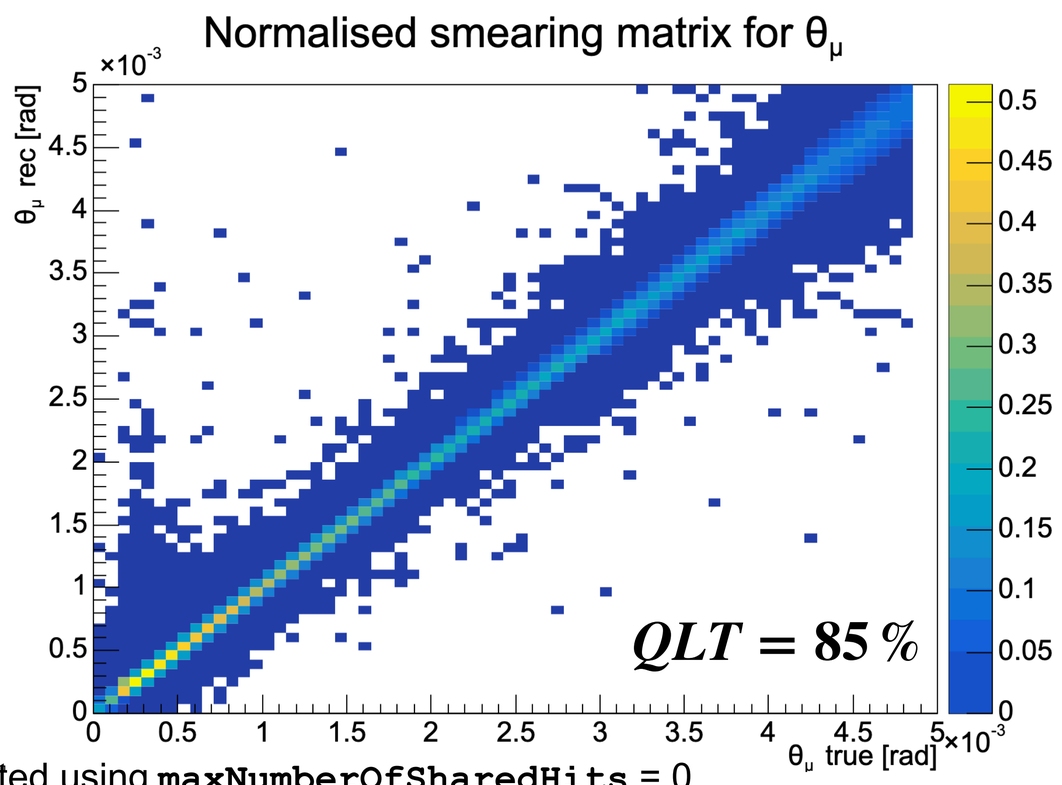




What is the Vertical Strip?

- As expected, since these events have low fraction of shared hits with linked track, if one raises the QLT cut the events on the strip are suppressed
- Of course this comes at the expense of efficiency (see later)

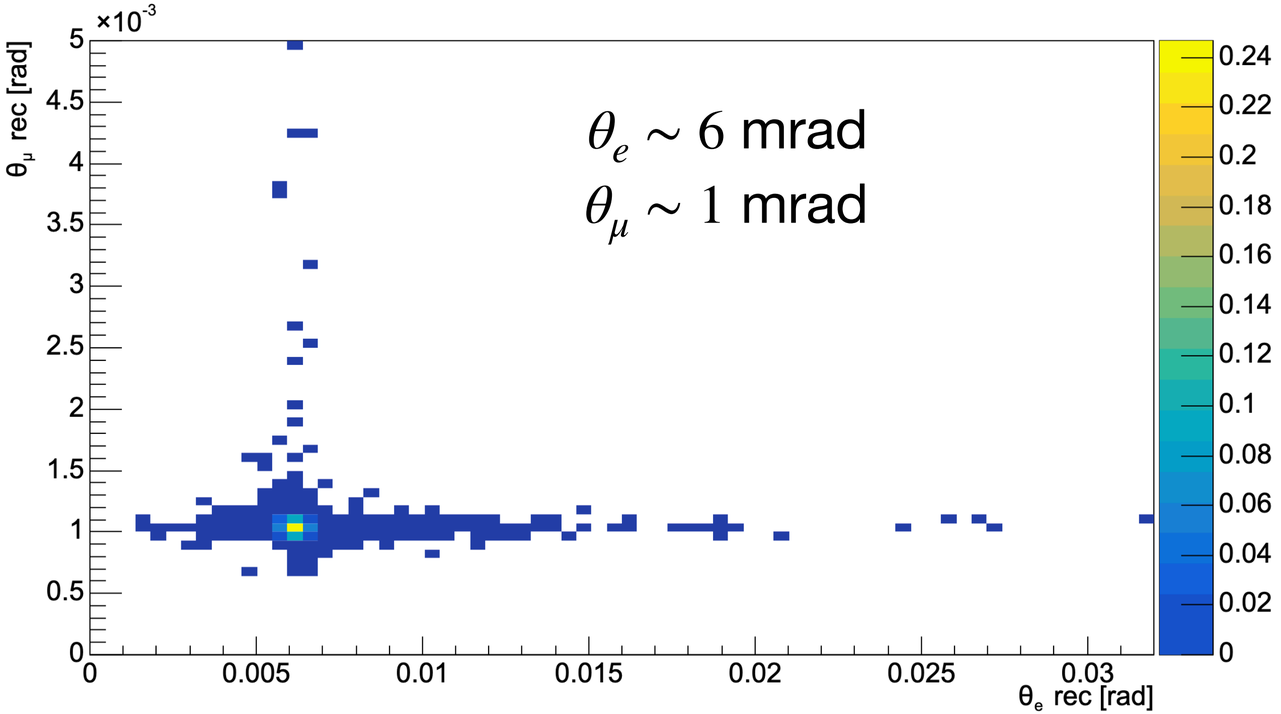


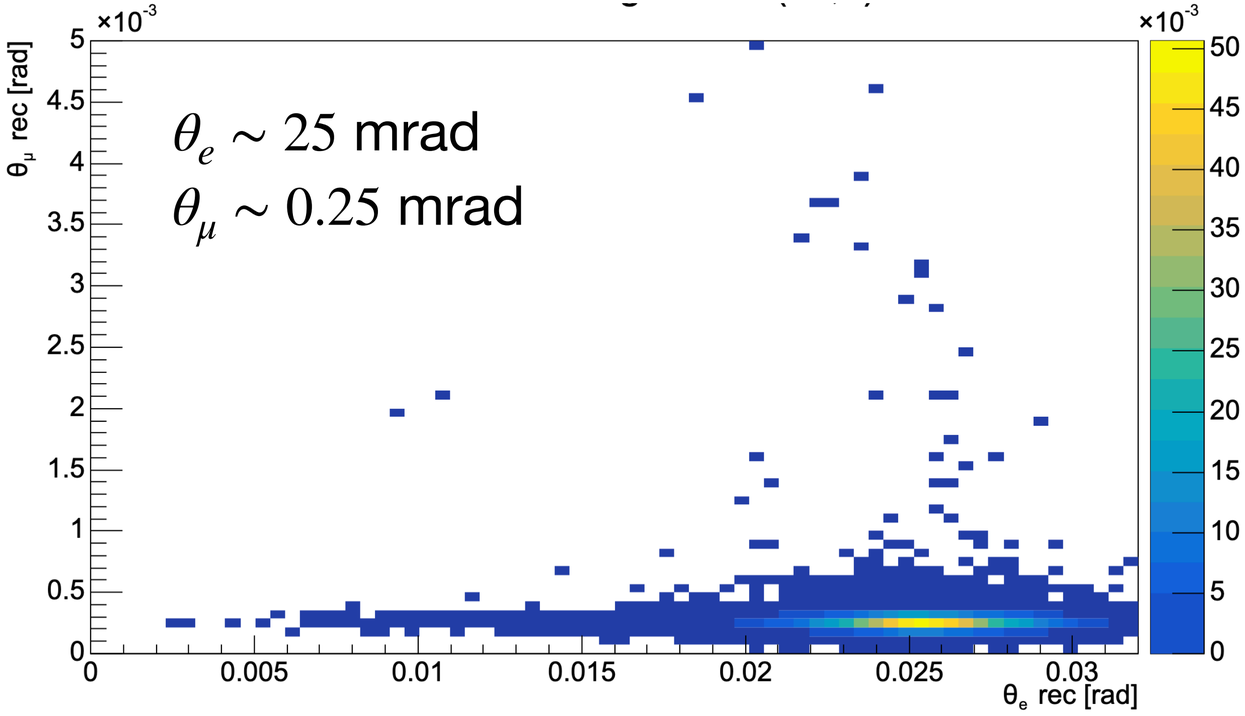


Plots generated using `maxNumberOfSharedHits` = 0

2D smearing Visualisation: *Mountain Plots*

- One way to visualise smearing in 2D is plotting the probability of reconstructing a fixed bin of $(\theta_e, \theta_\mu)_{true}$ in the space of $(\theta_e, \theta_\mu)_{reco}$



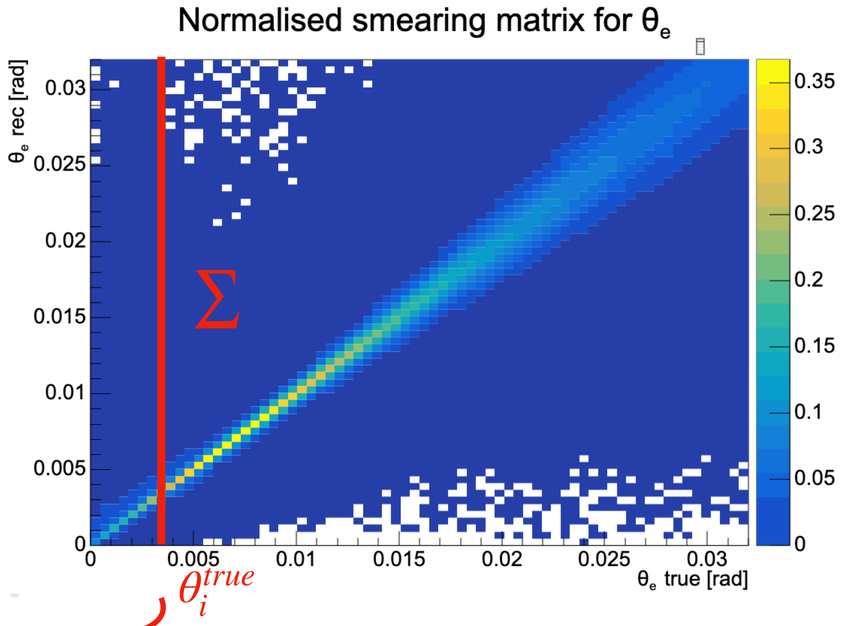


Plots generated using $QLT = 67\ \%$ and **`maxNumberOfSharedHits`** = 0

From Smearing Matrices to Efficiencies

- From the smearing matrices we obtain efficiencies by summing the probabilities of reconstructing an event belonging to a fixed bin of $\theta_{e/\mu}^{true}$ in any of the possible reconstructed bins $\theta_{e/\mu}^{reco}$

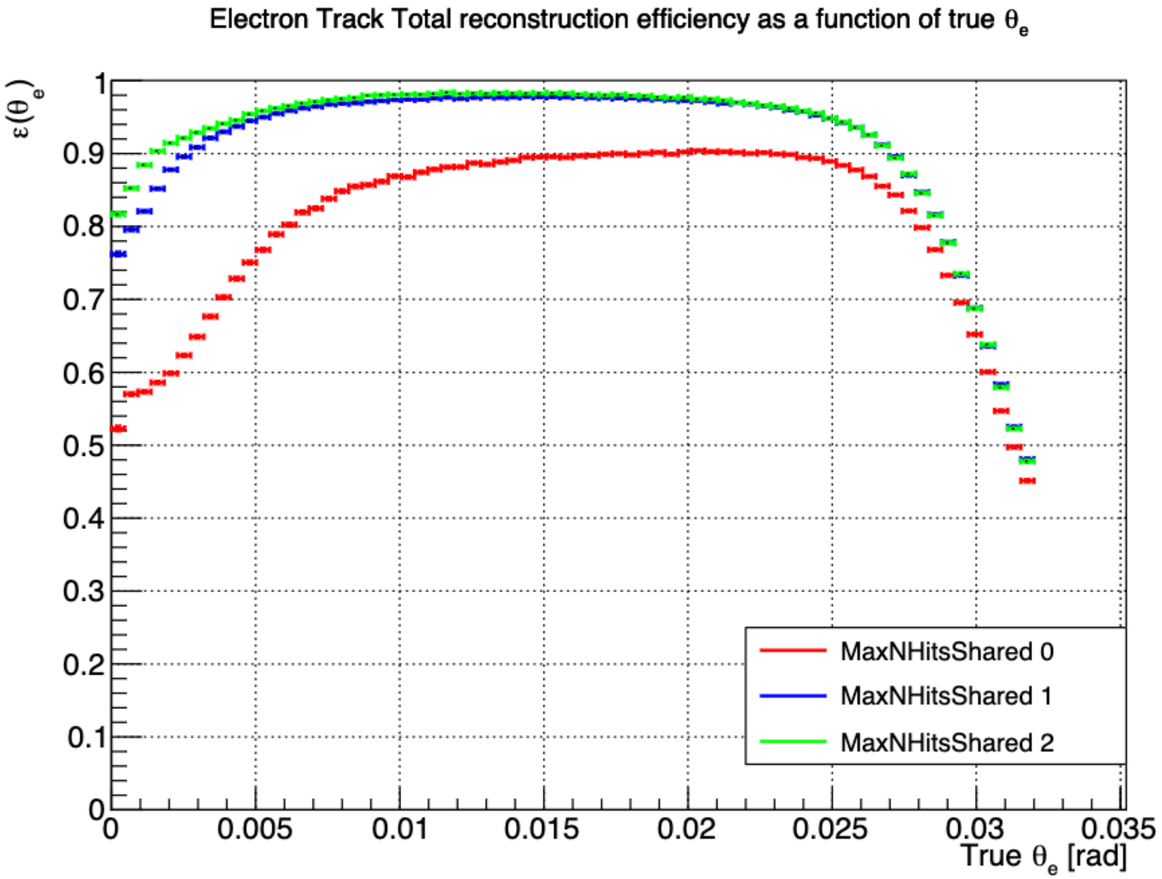
$$\varepsilon(\theta_i^{true}) = \sum_j P(\theta_j^{reco} \mid \theta_i^{true})$$



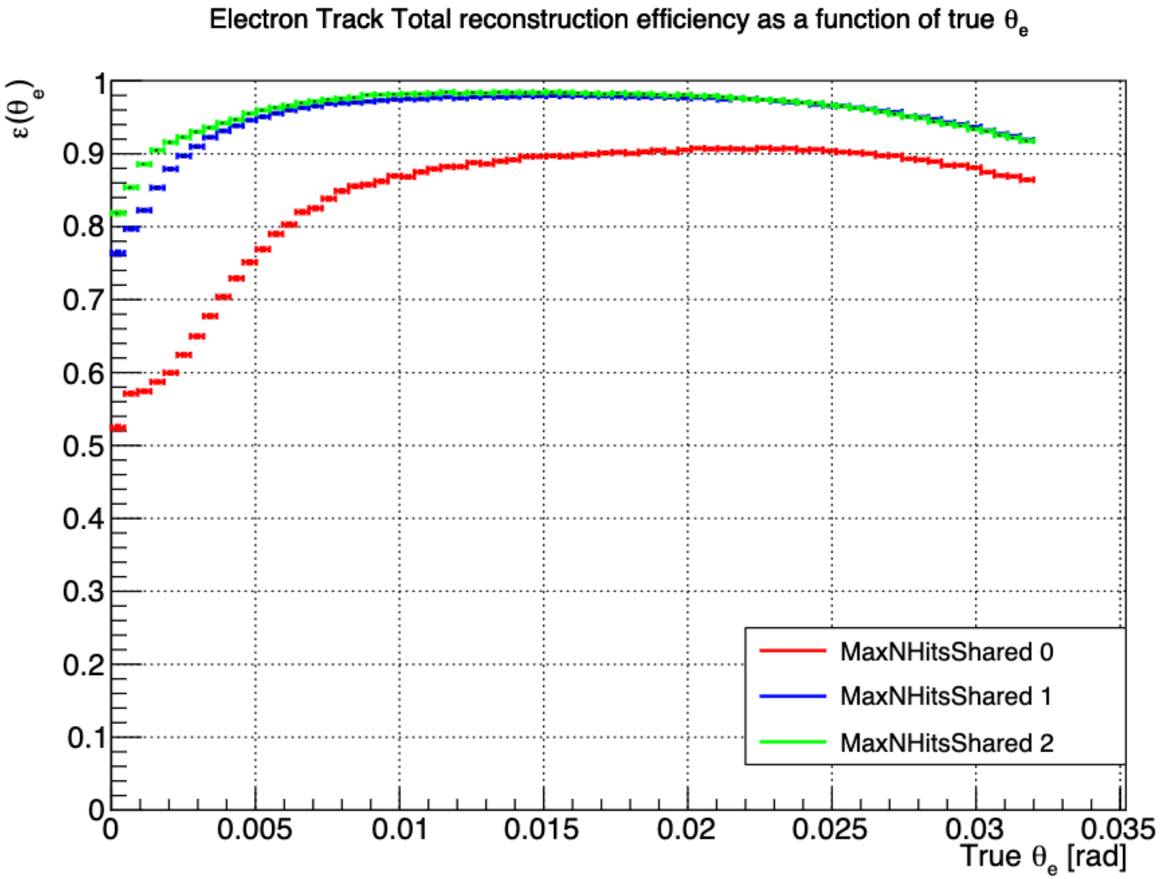
- In the 2D case this is done by unfolding the 2-dimensional $(\theta_e, \theta_\mu)_{true}$ in a single 1D object, and treating this flattened 1D histogram as detailed above. Finally, one wraps up the resulting 1D efficiency histogram to obtain a 2D efficiency histogram

- The overflow of the θ^{reco} histogram may or may not be considered in the efficiency

Global Track Efficiency $\varepsilon_{trk,e}^{glb}$ as a function of θ_e



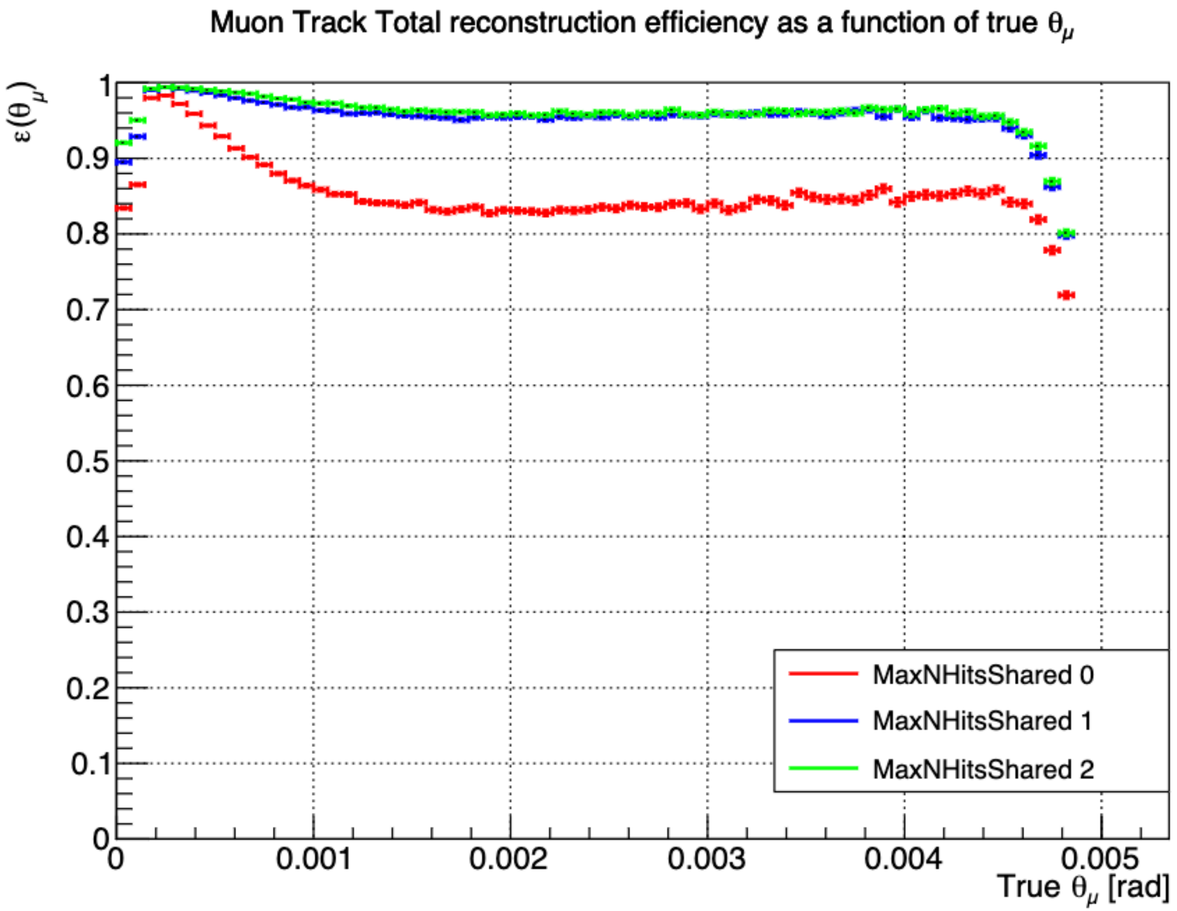
No overflow



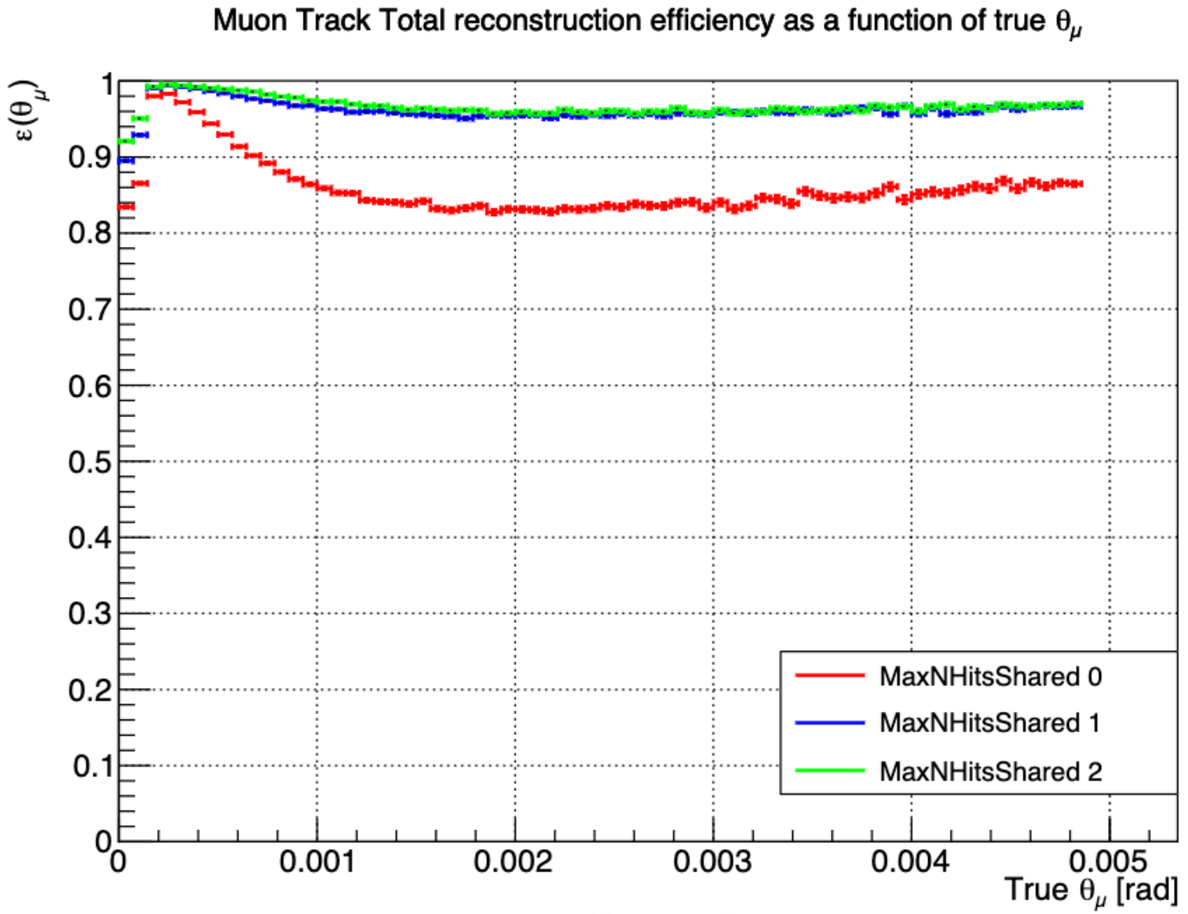
With Overflow

Plots generated using $QLT = 67\ \%$

Global Track Efficiency $\varepsilon^{glb}_{trk,\mu}$ as a function of θ_μ



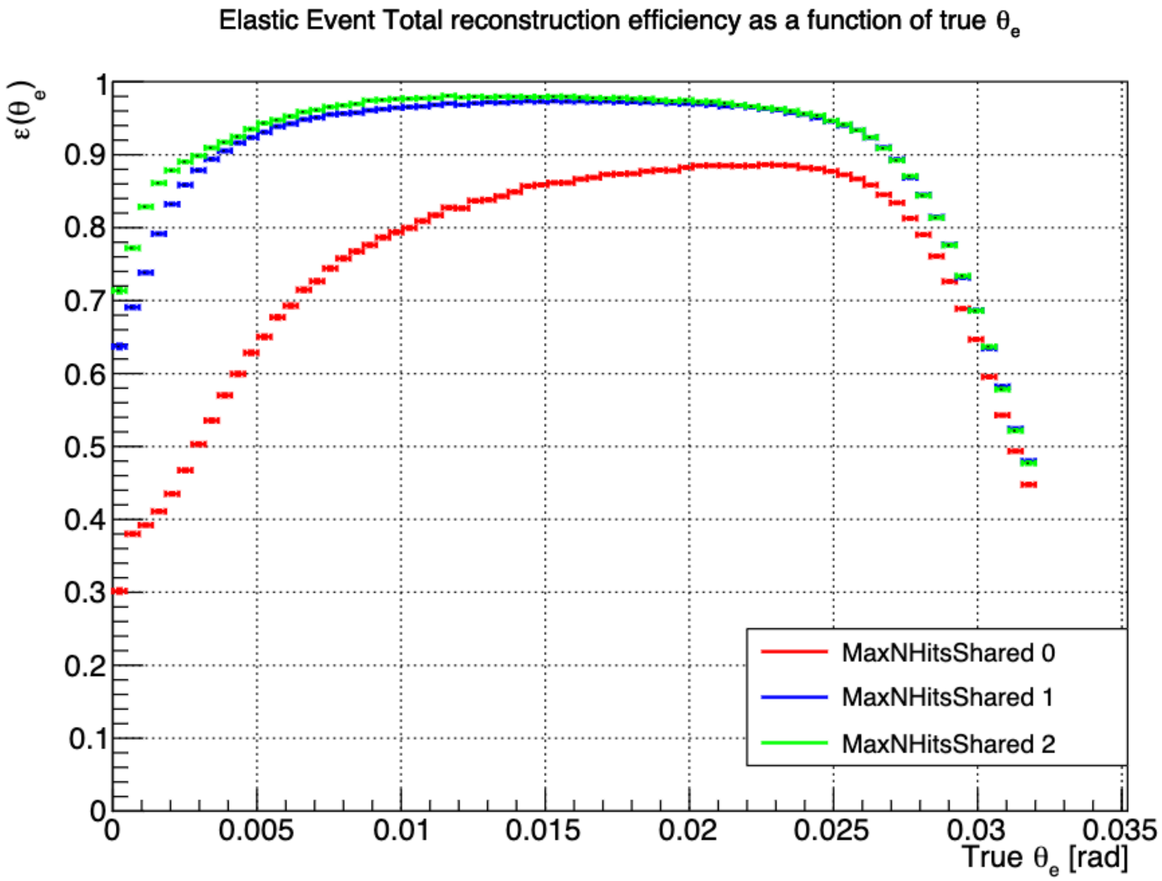
No overflow



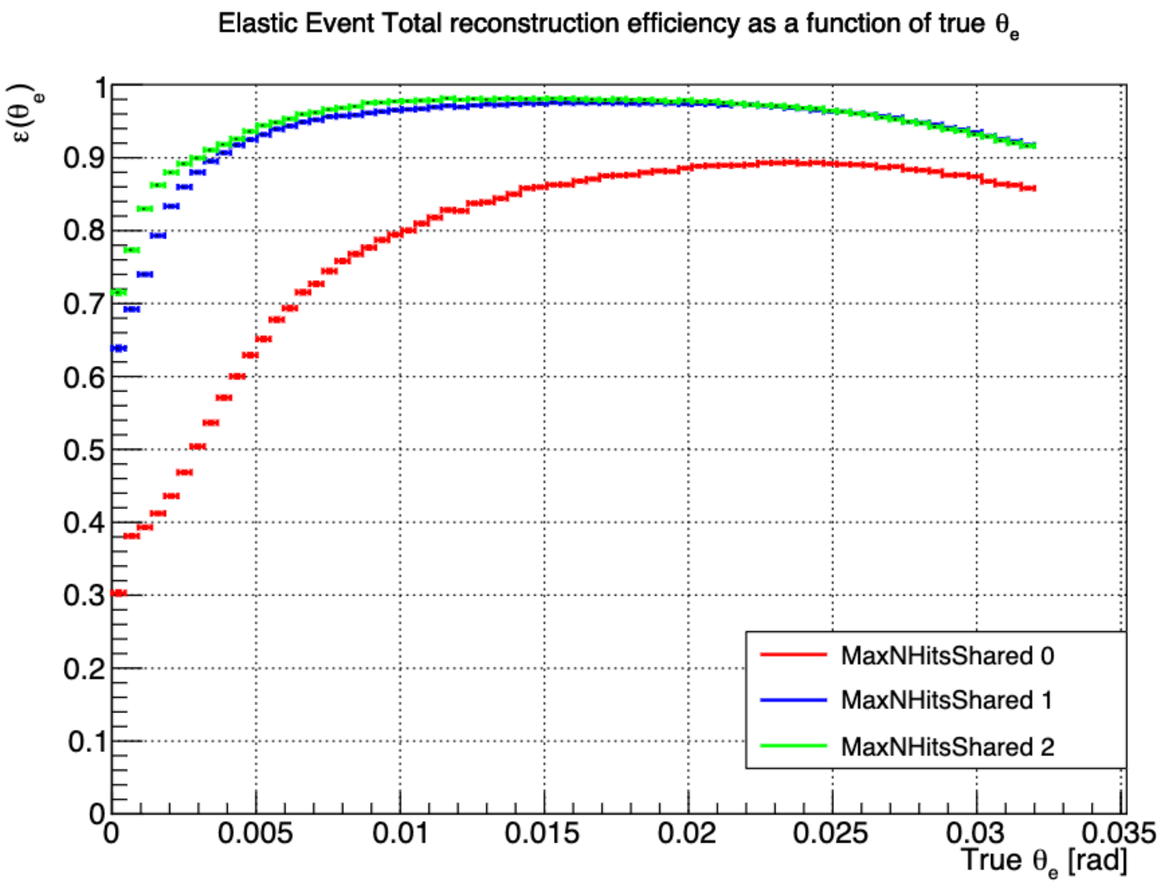
With Overflow

Plots generated using $QLT = 67\ \%$

Global Event Efficiency ε_{ev}^{glb} as a function of θ_e



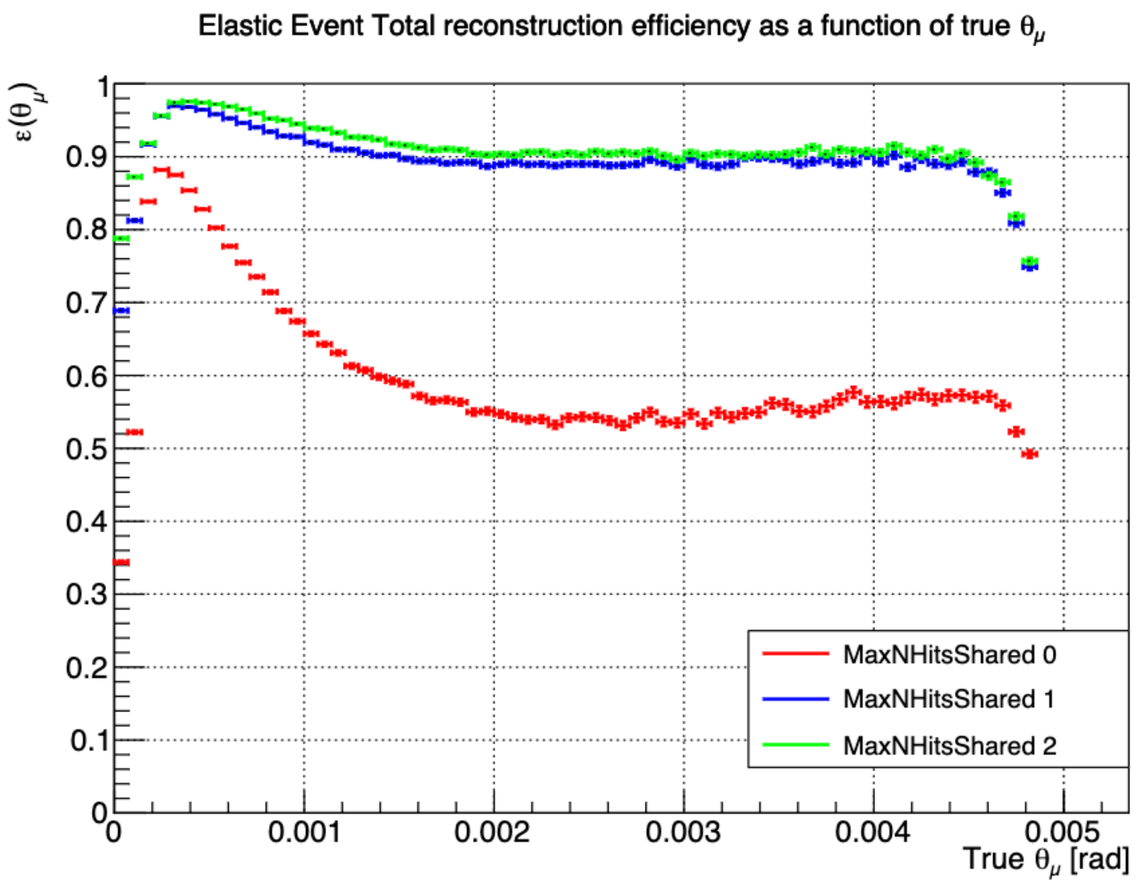
No overflow



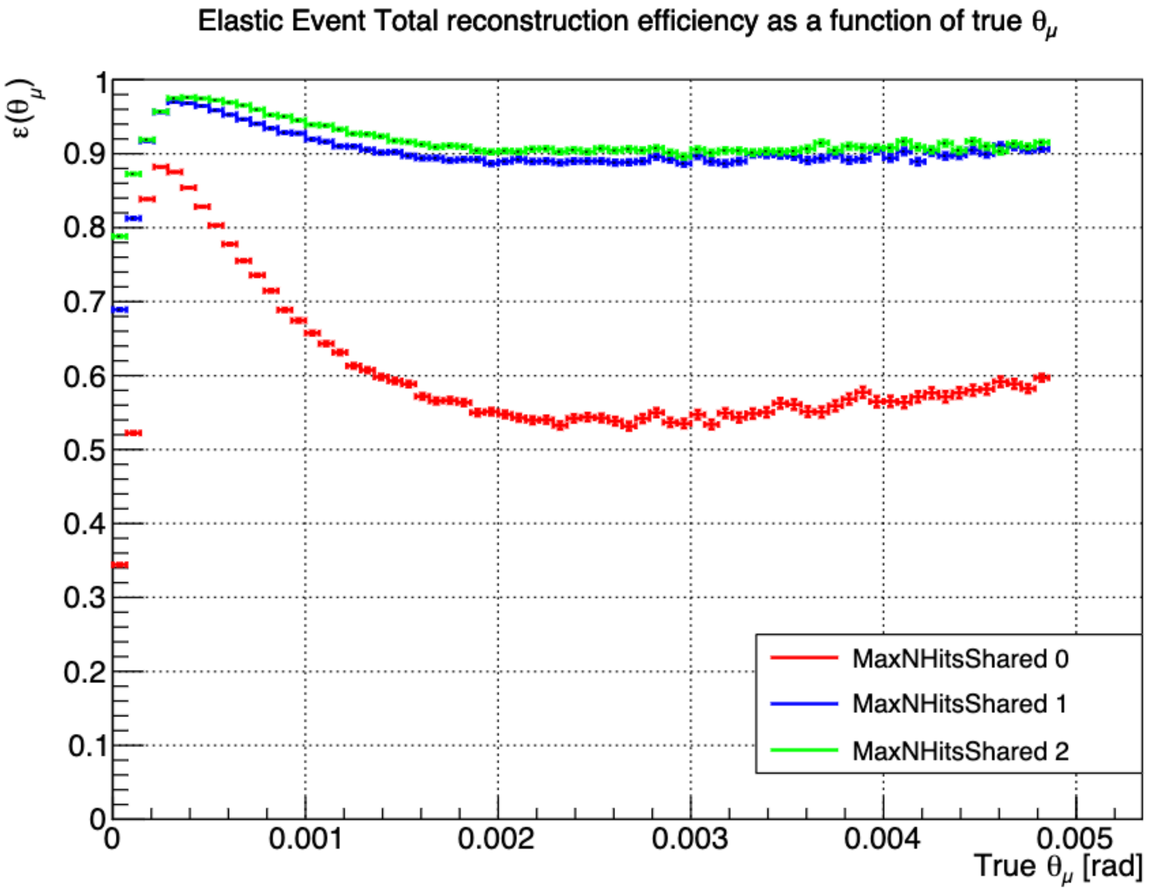
With Overflow

Plots generated using $QLT = 67\ \%$

Global Event Efficiency ε_{ev}^{glb} as a function of θ_μ



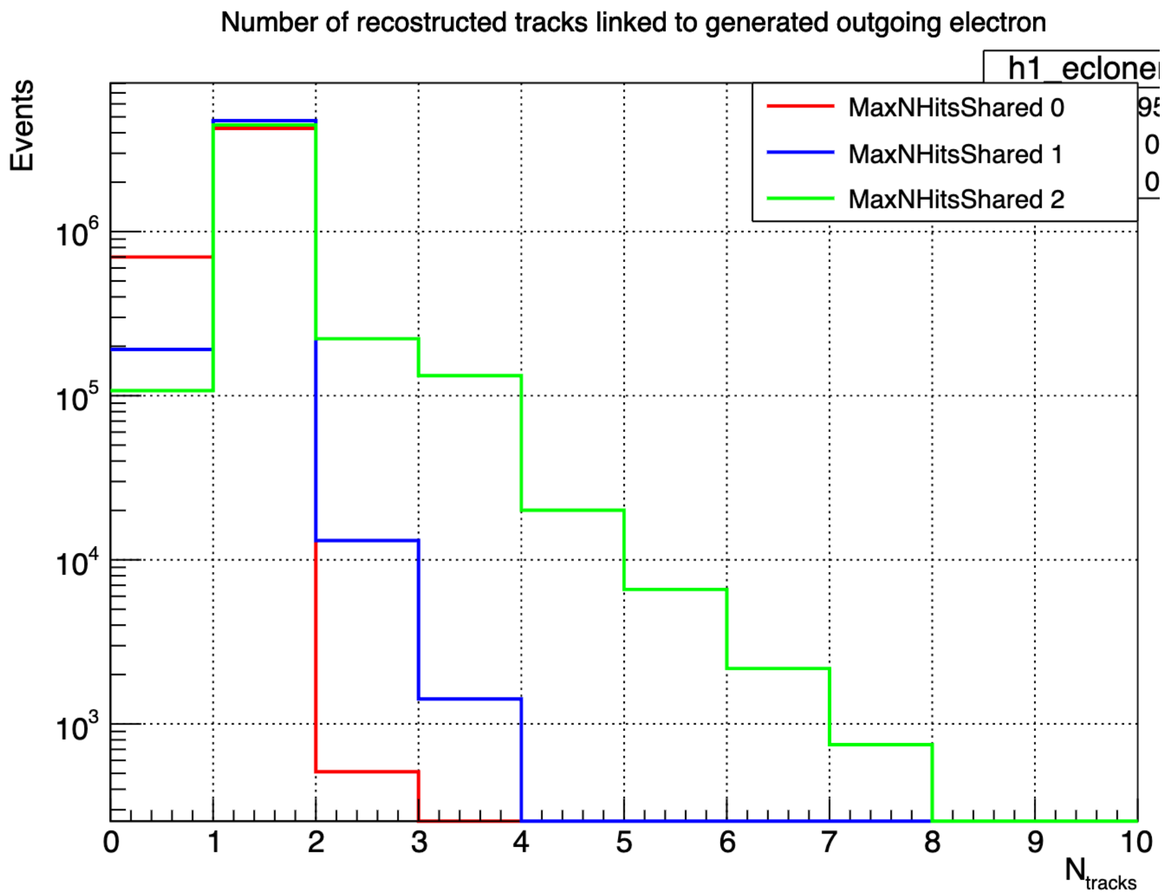
No overflow

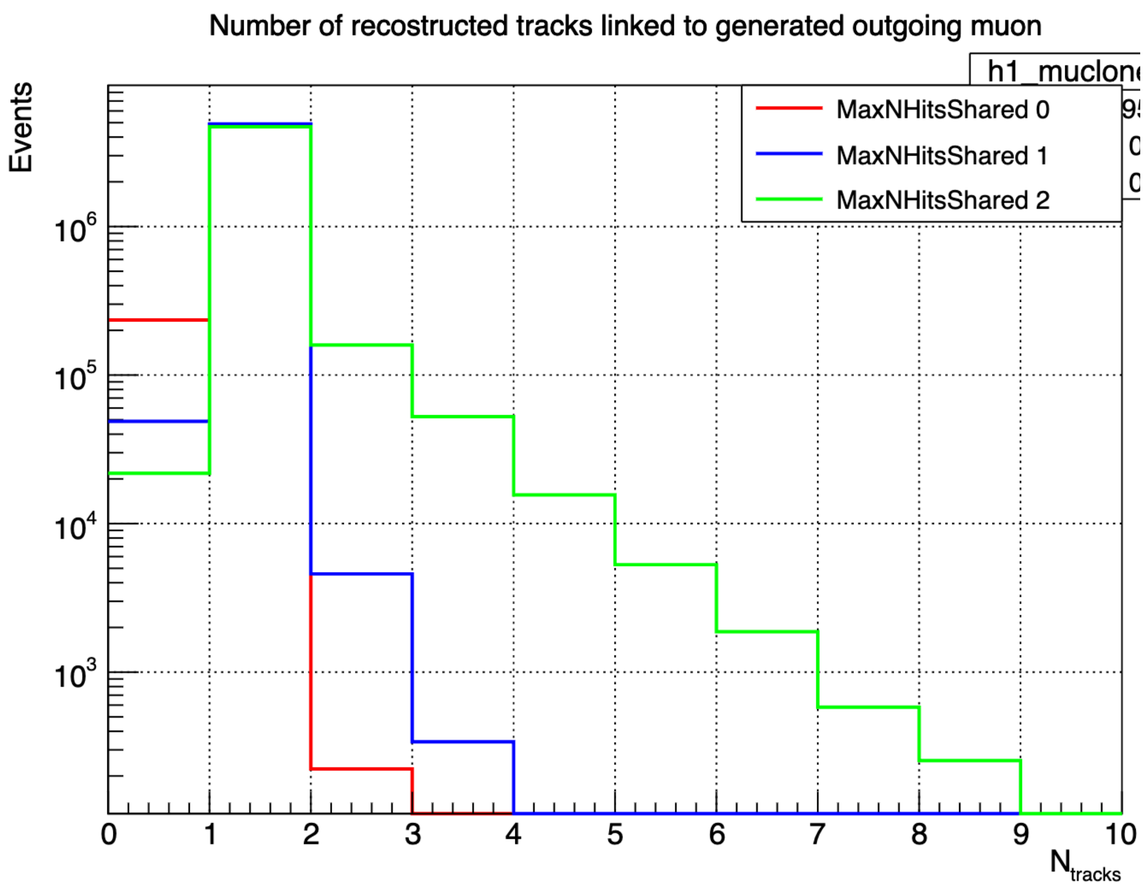


With Overflow

Plots generated using $QLT = 67\ \%$

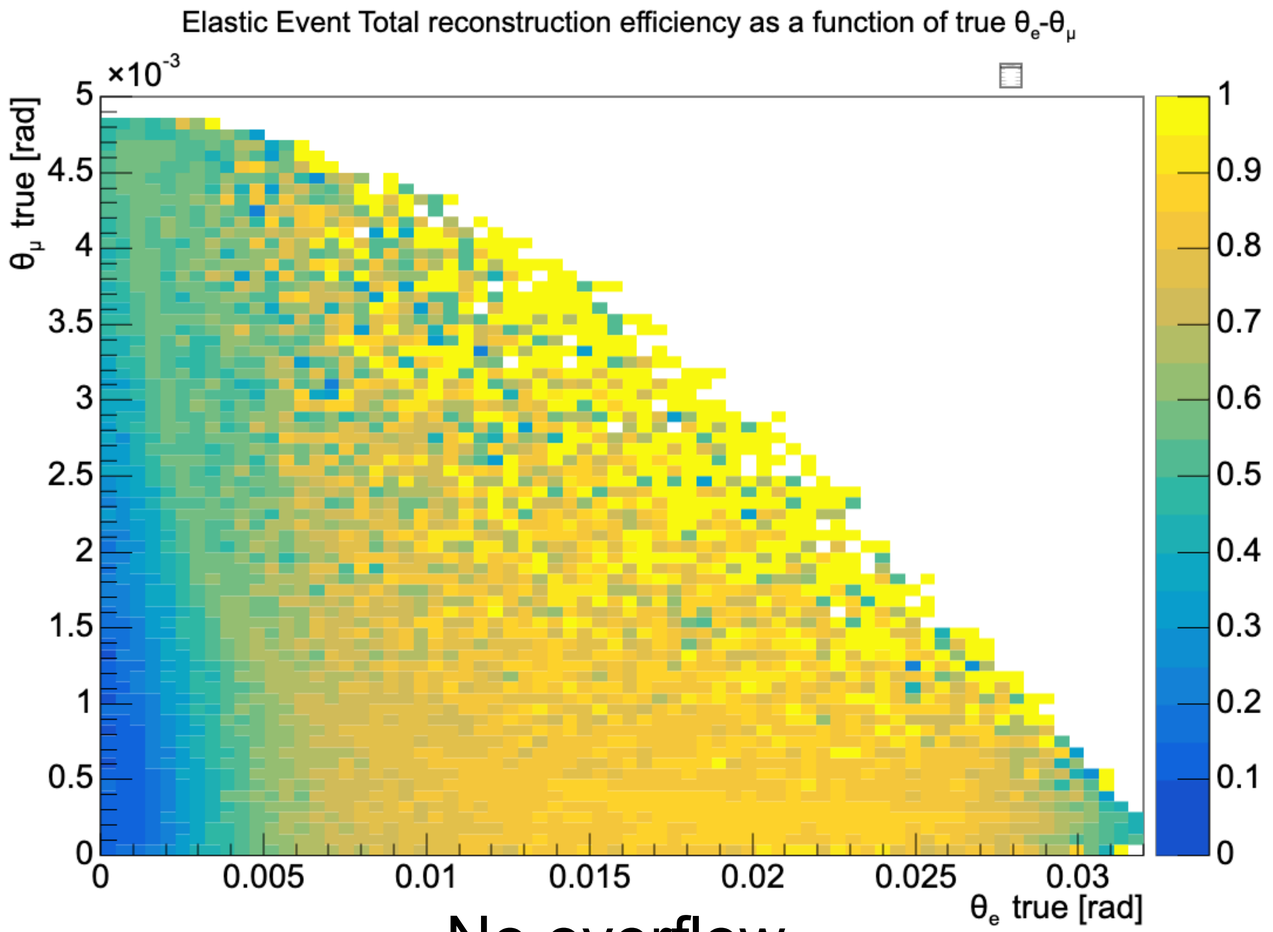
More Shared Hits = More Clones



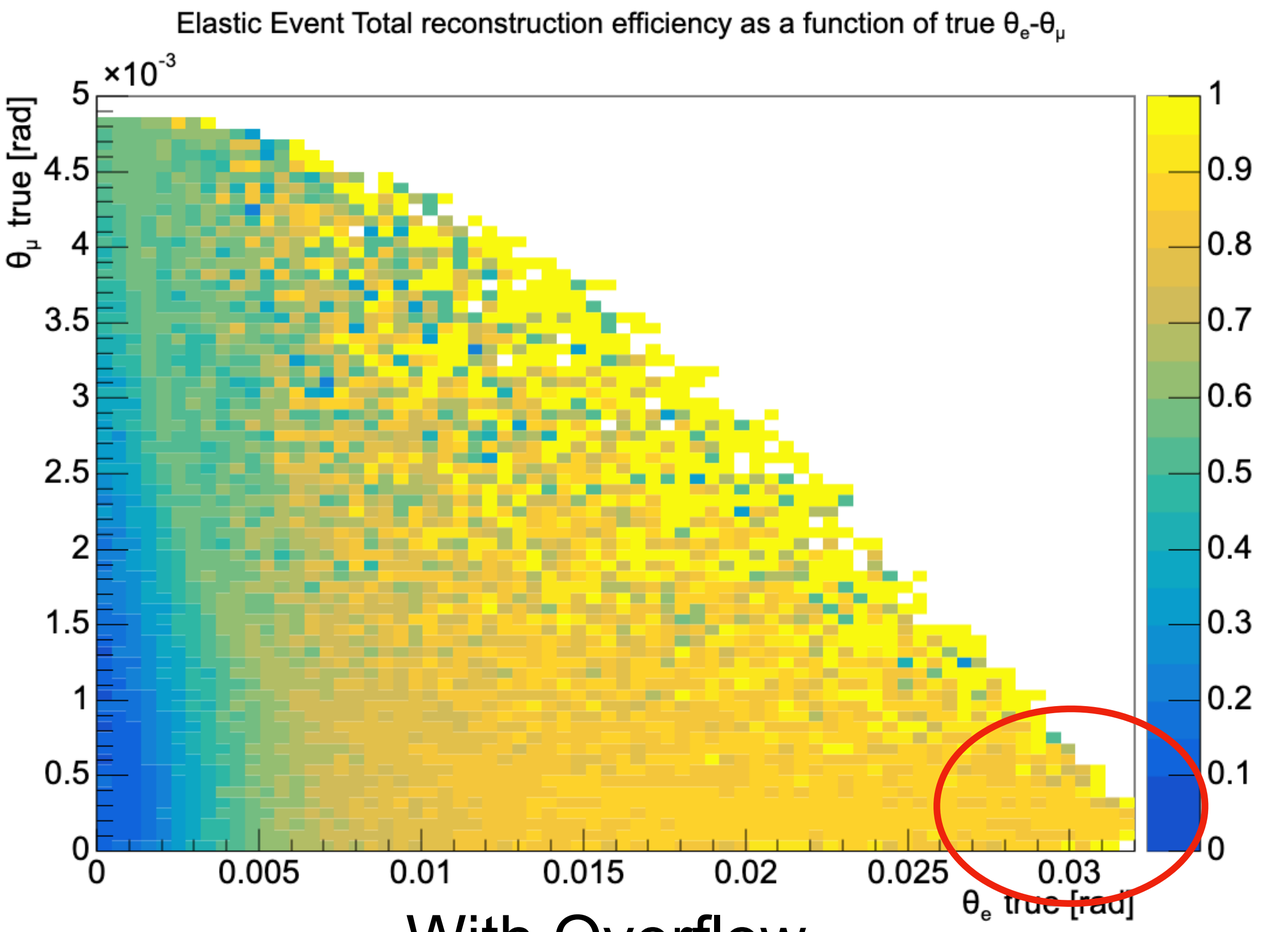


This is the number of tracks whose linked track is the MC truth outgoing electron and muon respectively

Global Event Efficiency ε_{ev}^{glb} as a function of (θ_e, θ_μ)



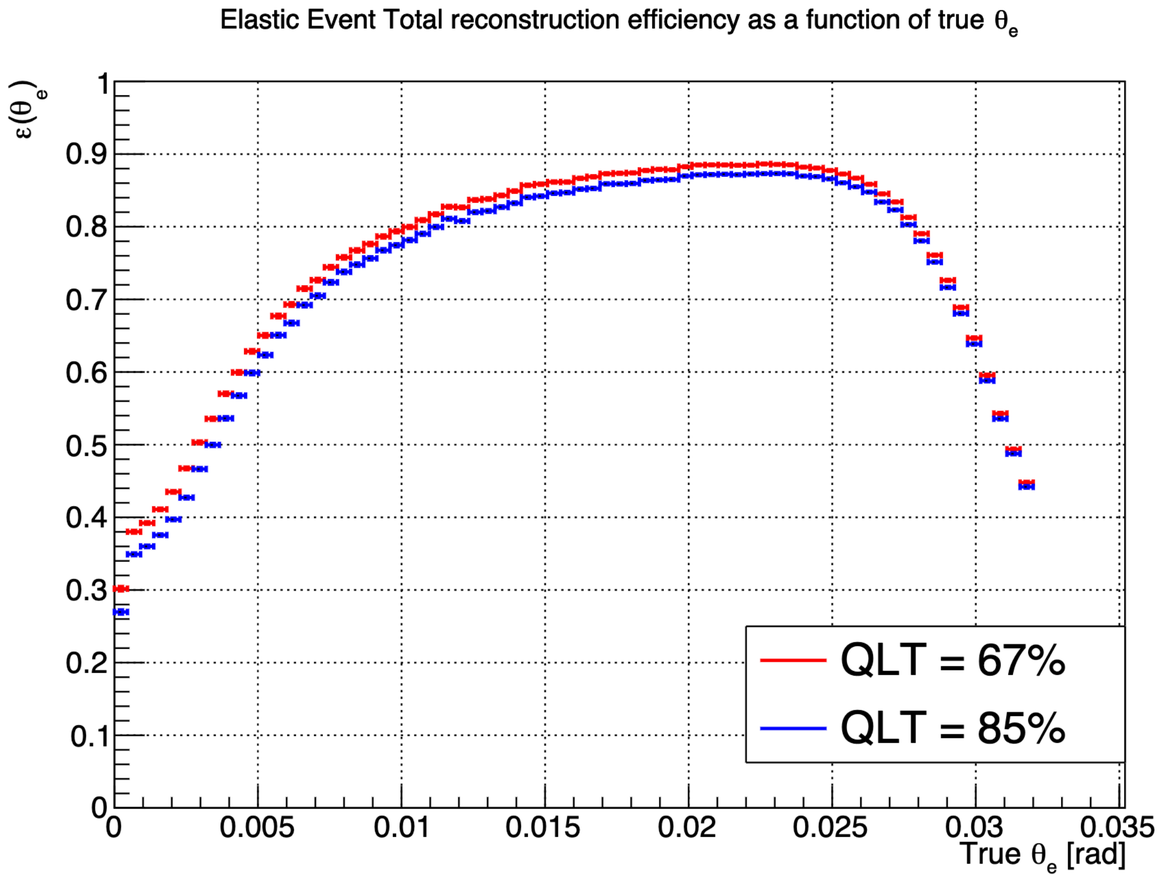
No overflow

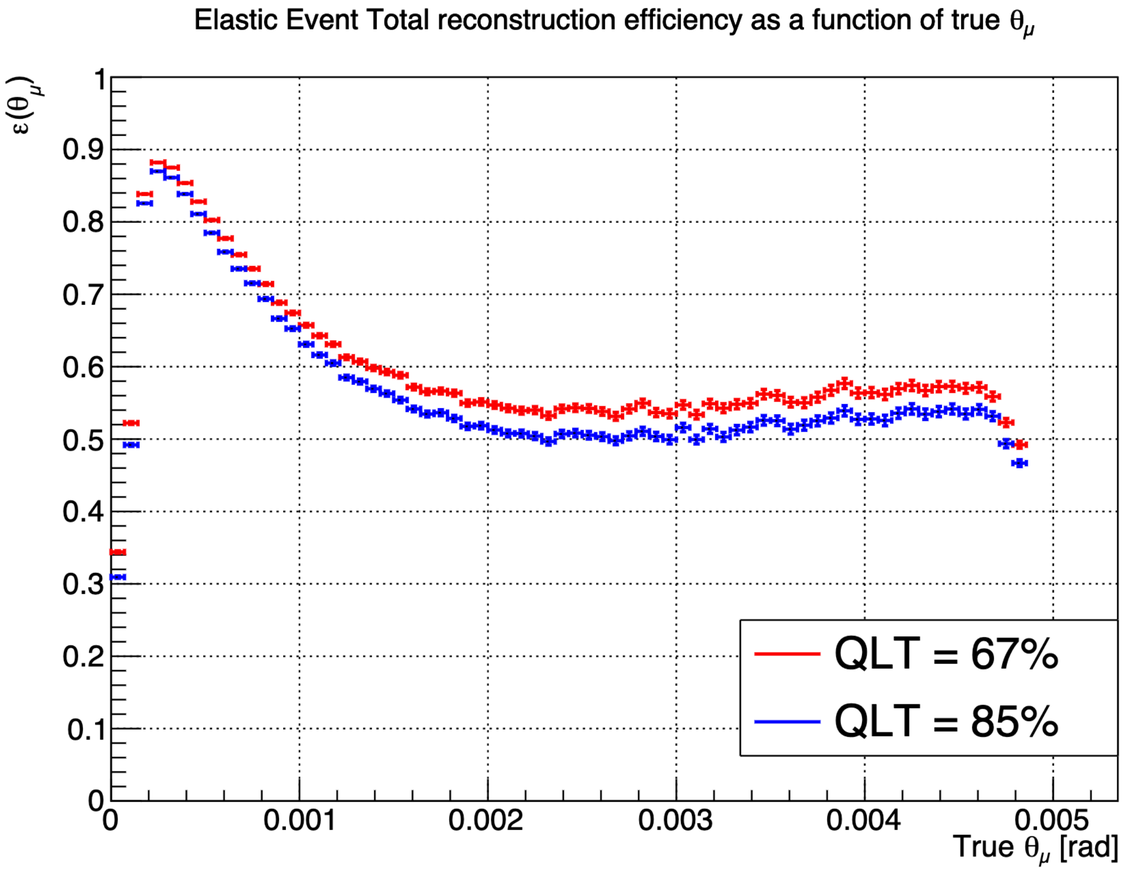


With Overflow

Plots generated using $QLT = 67\,\%$ and **`maxNumberOfSharedHits`** = 0

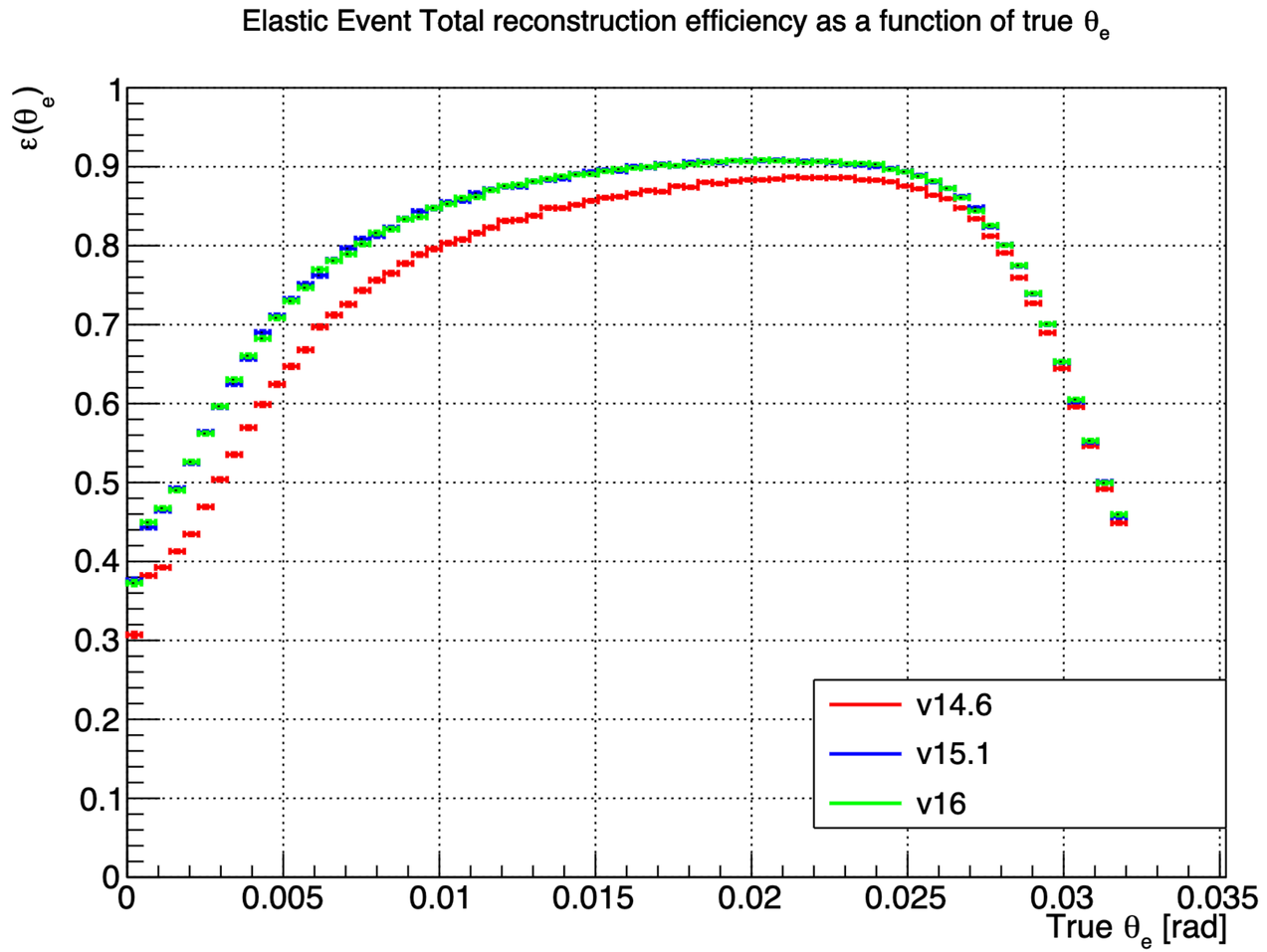
Global Event Efficiency ε_{ev}^{glb} with different QLT





Plots generated using **`maxNumberOfSharedHits`** = 0 and no overflow

Reconstructable Event Efficiency ε_{ev}^{rca} in different versions



Plot generated using $QLT = 67\,\%$ and no overflow and **`maxNumberOfSharedHits`** = 0

Reconstructable Efficiency ε^{rca} Studies

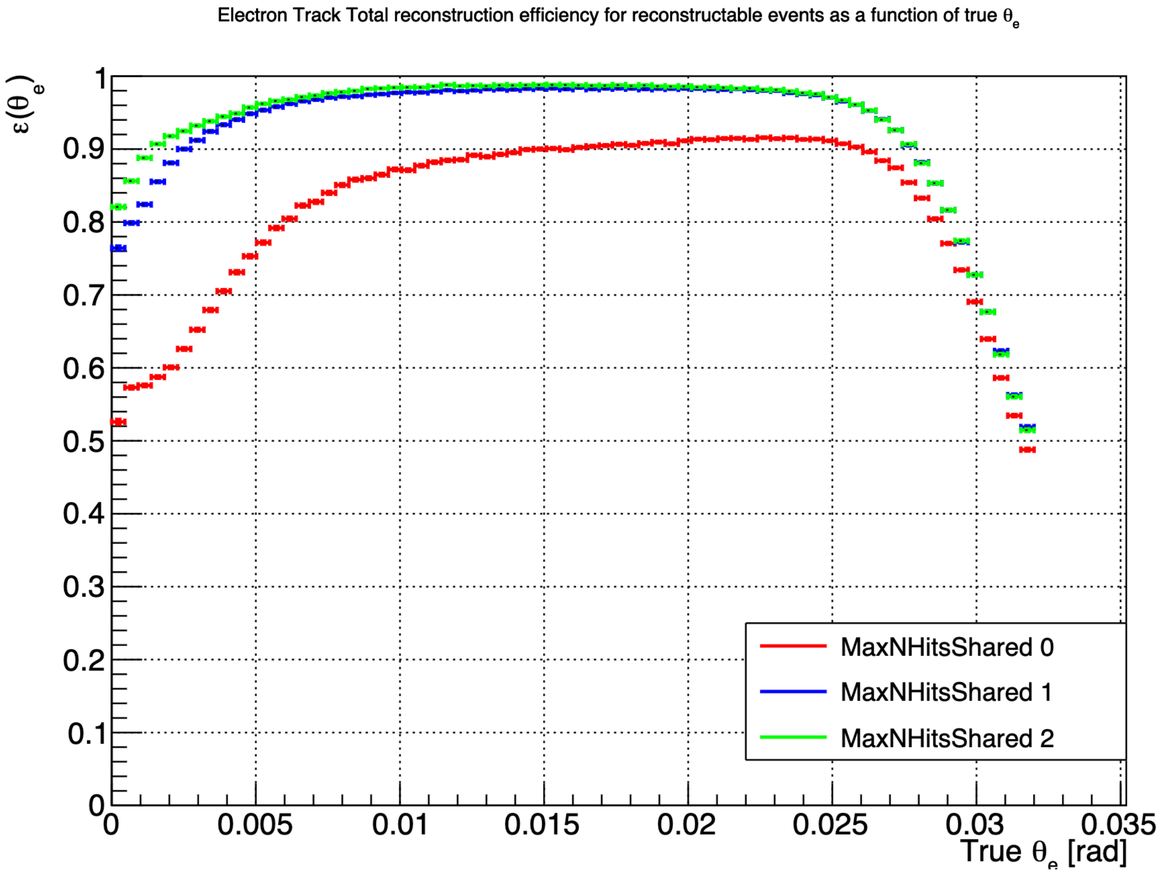
Definition of Reconstructable Efficiency ε^{rca}

- The reconstructable elastic event (ev) efficiency and the two reconstructable outgoing track (trk) efficiencies are determined in the following way:

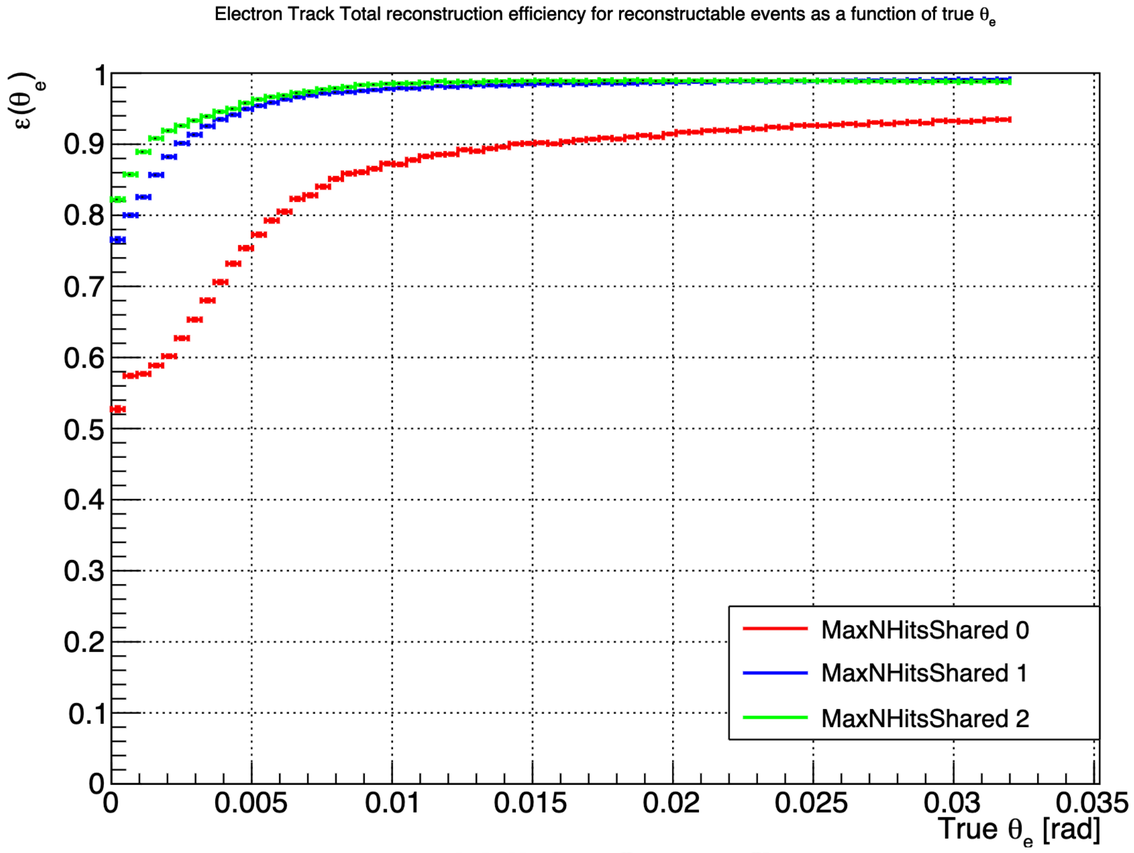
$$\varepsilon_{ev}^{rca} = \frac{N\left[GEN \cap RCA \cap ALLREC\right]}{N\left[GEN \cap RCA\right]} \qquad \varepsilon_{trk}^{rca} = \frac{N\left[GEN \cap RCA \cap (ALLREC \cup TRKREC)\right]}{N\left[GEN \cap RCA\right]} \geq \varepsilon_{ev}^{rca}$$

- GEN : Incoming simulated μ_{in}^{+} hits target and both outgoing μ_{out}^{+} and e_{out}^{-} are generated in Monte Carlo
- $ALLREC$: All 3 tracks of the elastic event are reconstructed
- $TRKREC$: The incoming μ_{in}^{+} and the track under test are both reconstructed
- RCA : ***Reconstructable*** event $\equiv$ All 3 generated tracks have ≥ 5 hits (2X, 2Y, ≥ 1 stereo) in the respective station in the GEANT4 simulation + the μ_{in}^{+} track is reconstructed with 6 hits and a $\chi^2/ndof < 2$

Reconstructable Track Efficiency $\varepsilon_{trk,e}^{rca}$ as a function of θ_e



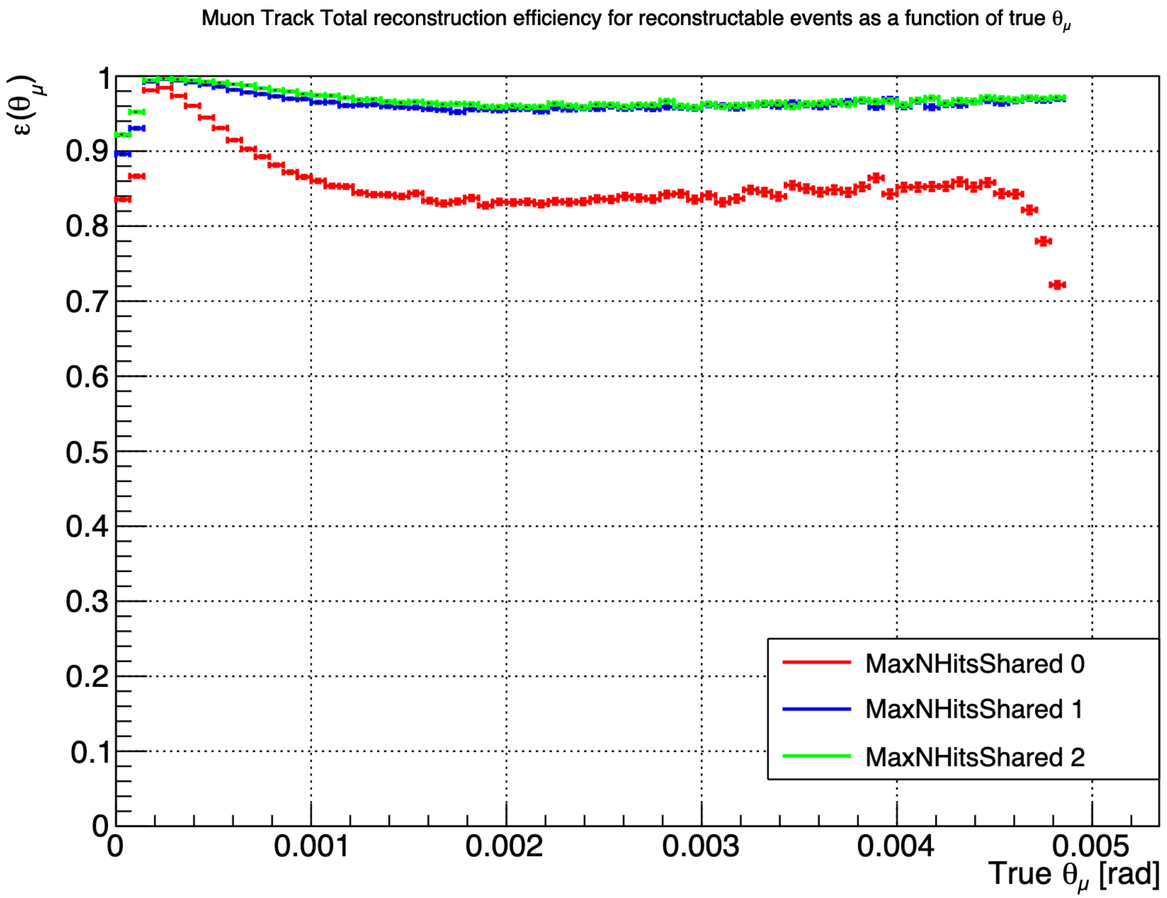
No overflow



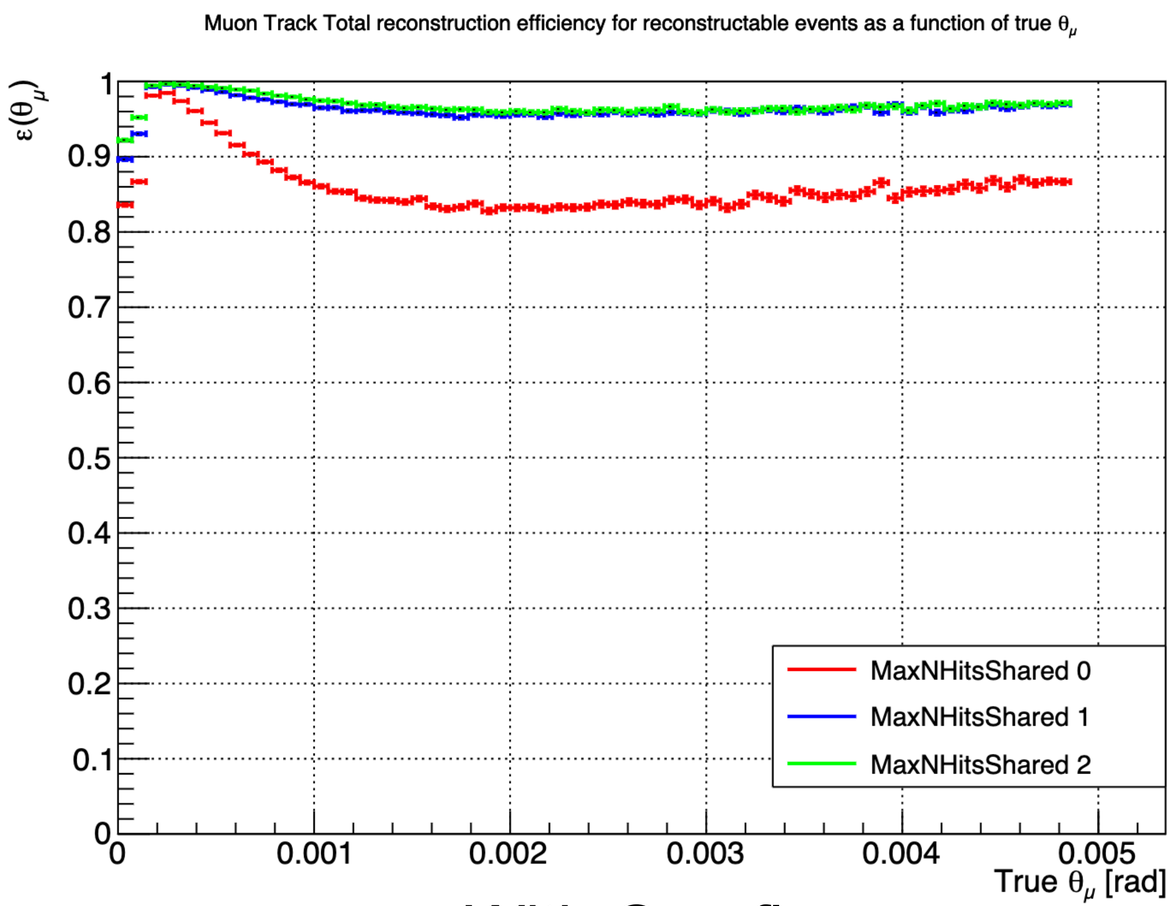
With Overflow

Plots generated using $QLT = 67\ \%$

Reconstructable Track Efficiency $\varepsilon_{trk,\mu}^{rca}$ as a function of θ_μ



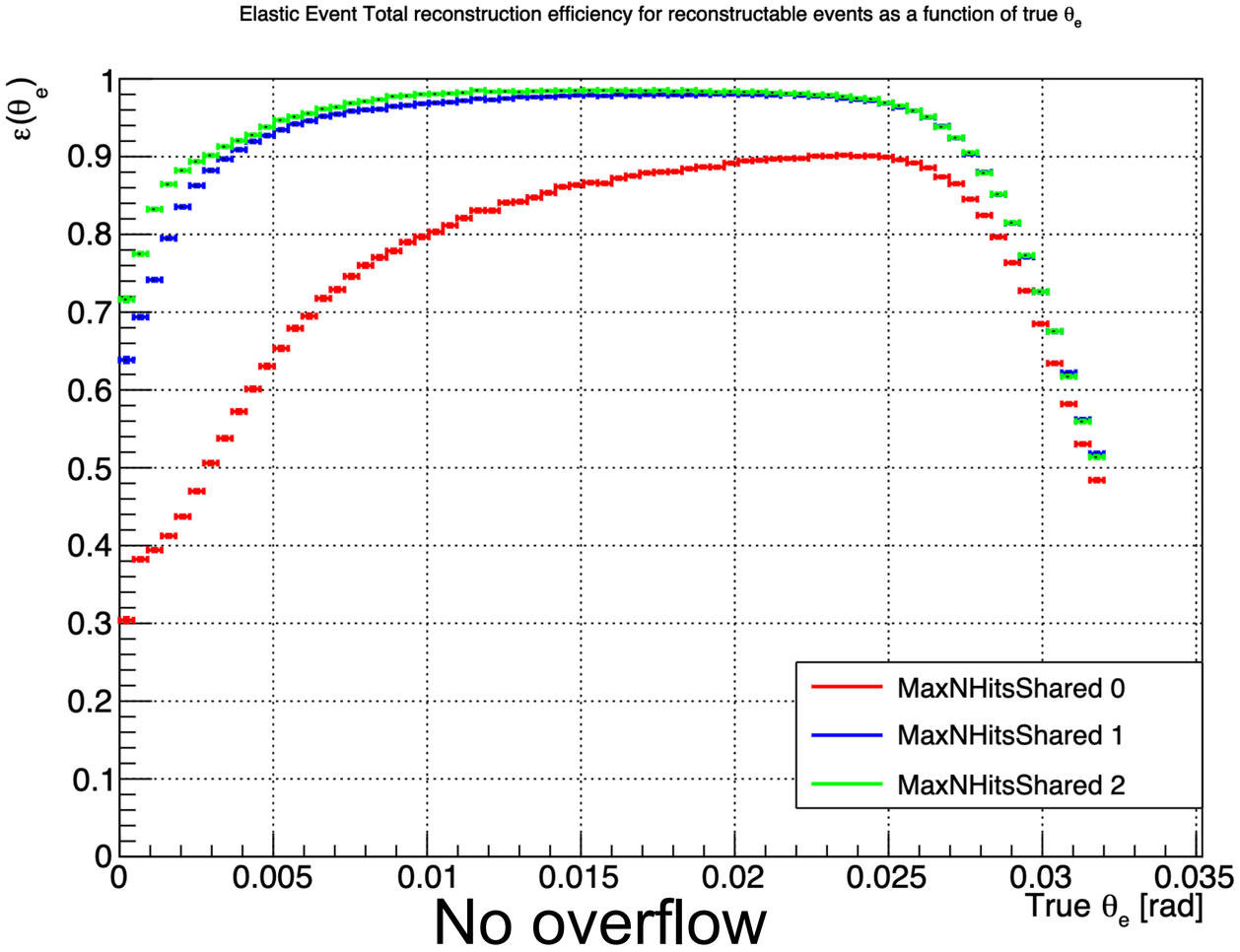
No overflow

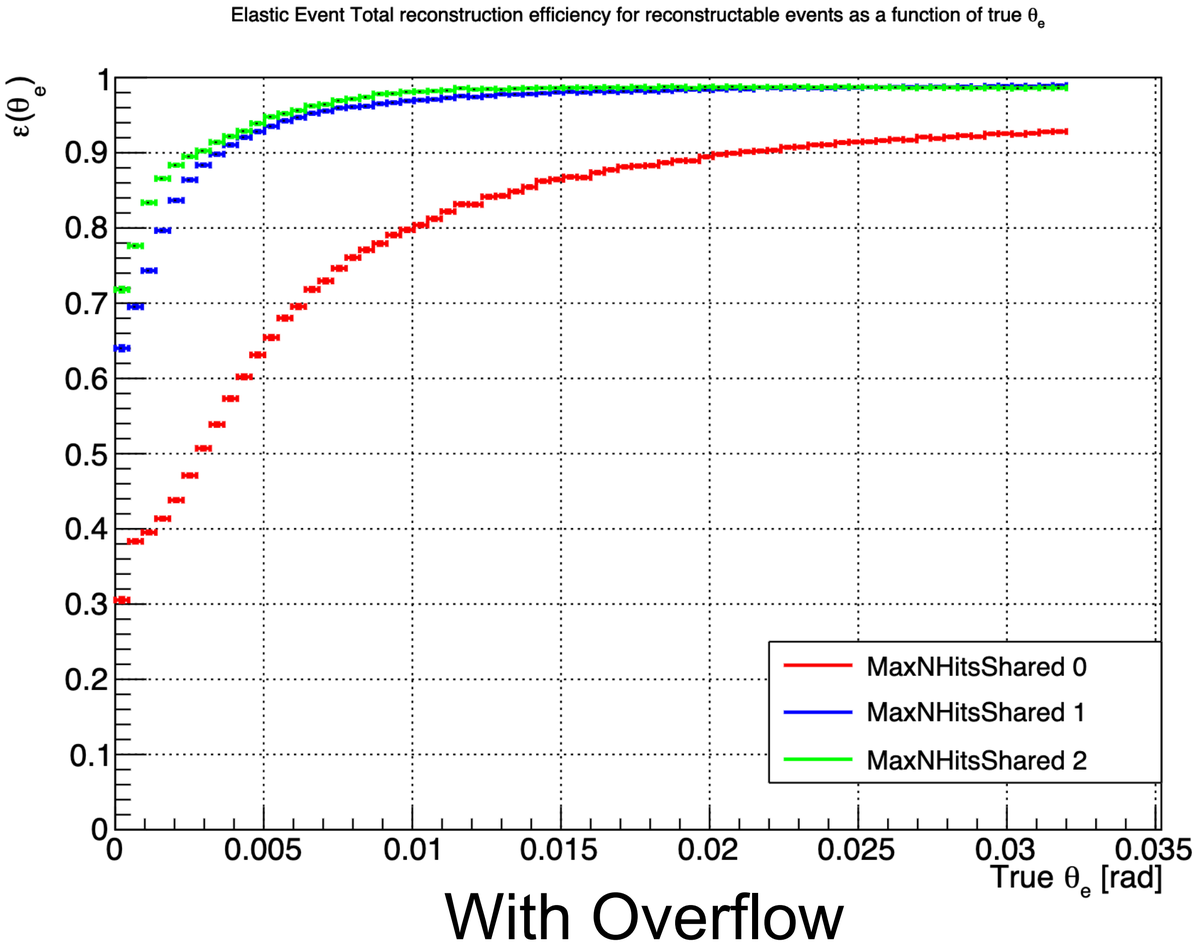


With Overflow

Plots generated using $QLT = 67\ \%$

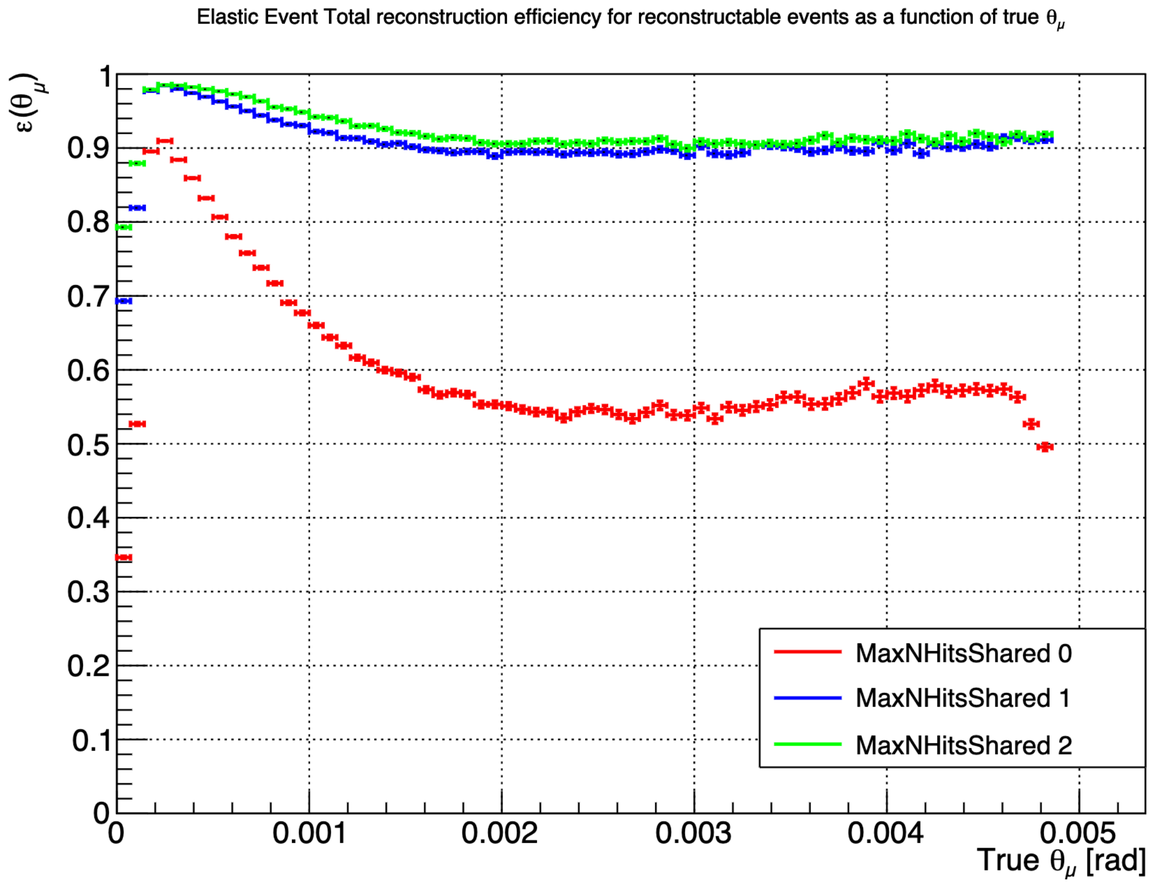
Reconstructable Event Efficiency ε_{ev}^{rca} as a function of θ_e



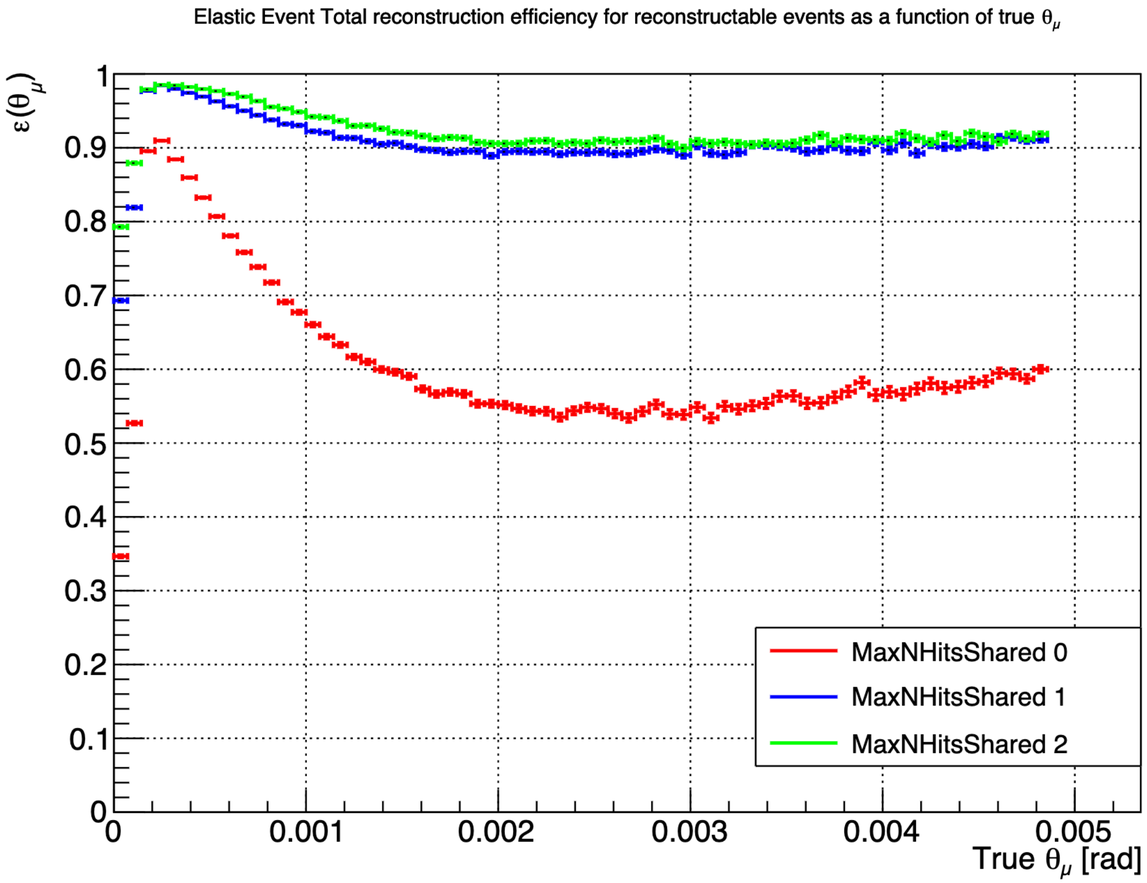


Plots generated using $QLT = 67\ \%$

Reconstructable Event Efficiency ε_{ev}^{rca} as a function of θ_μ



No overflow

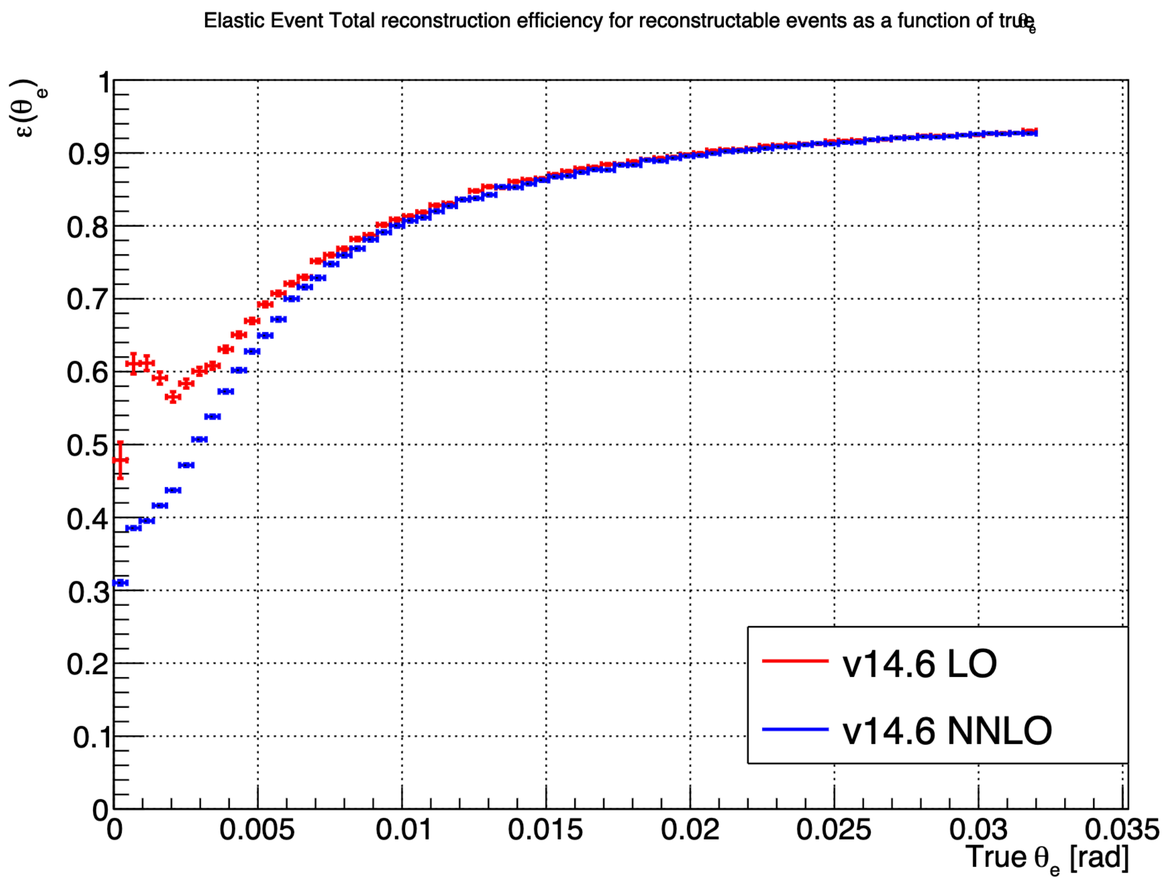


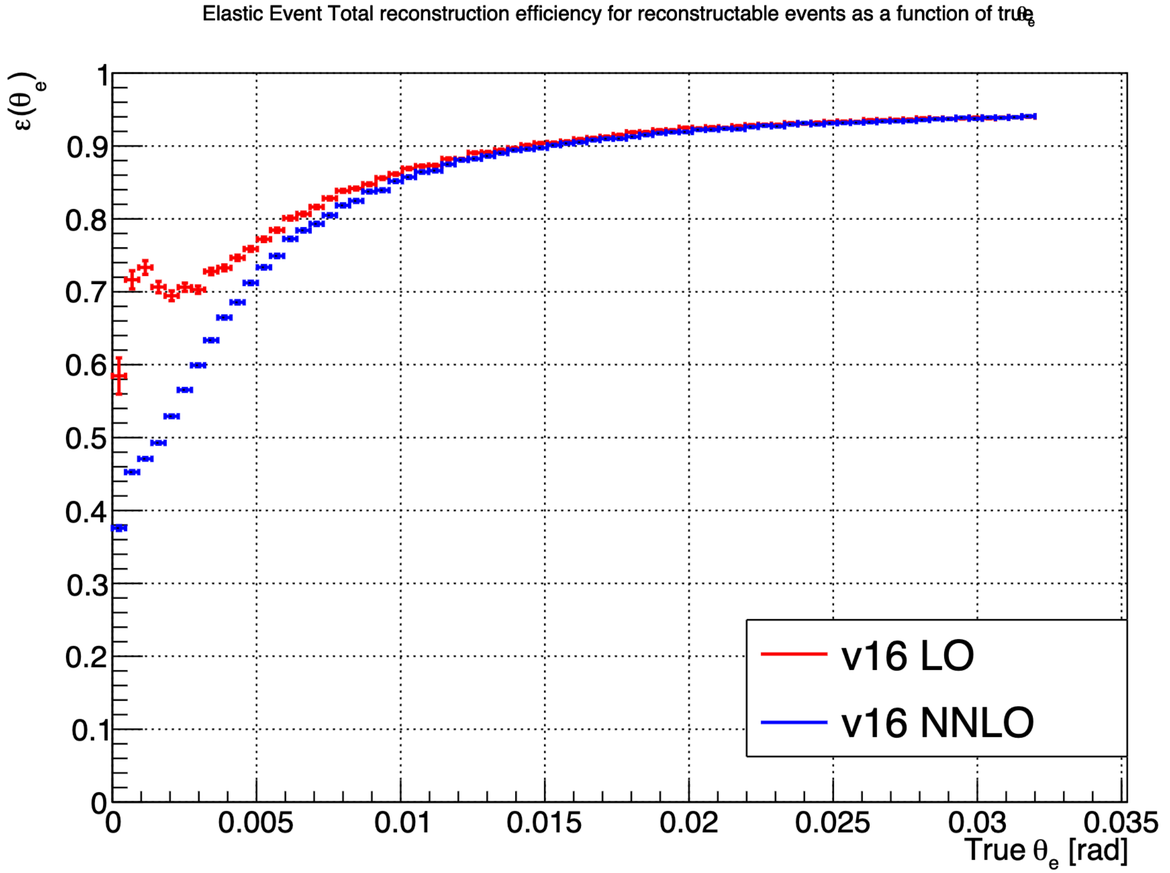
With Overflow

Plots generated using $QLT = 67\,\%$

LO vs NNLO and Comparison with Eugenia

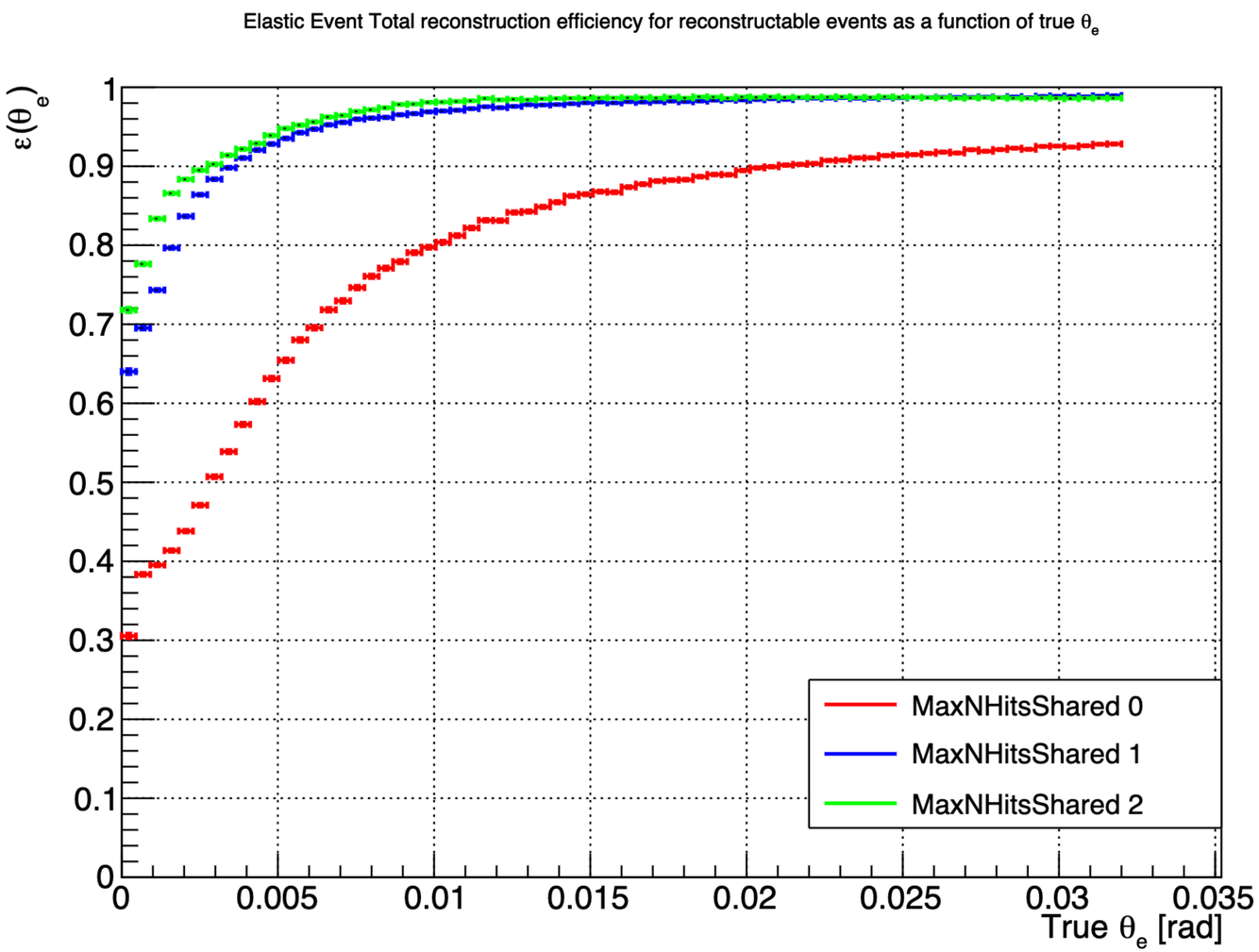
NNLO vs LO

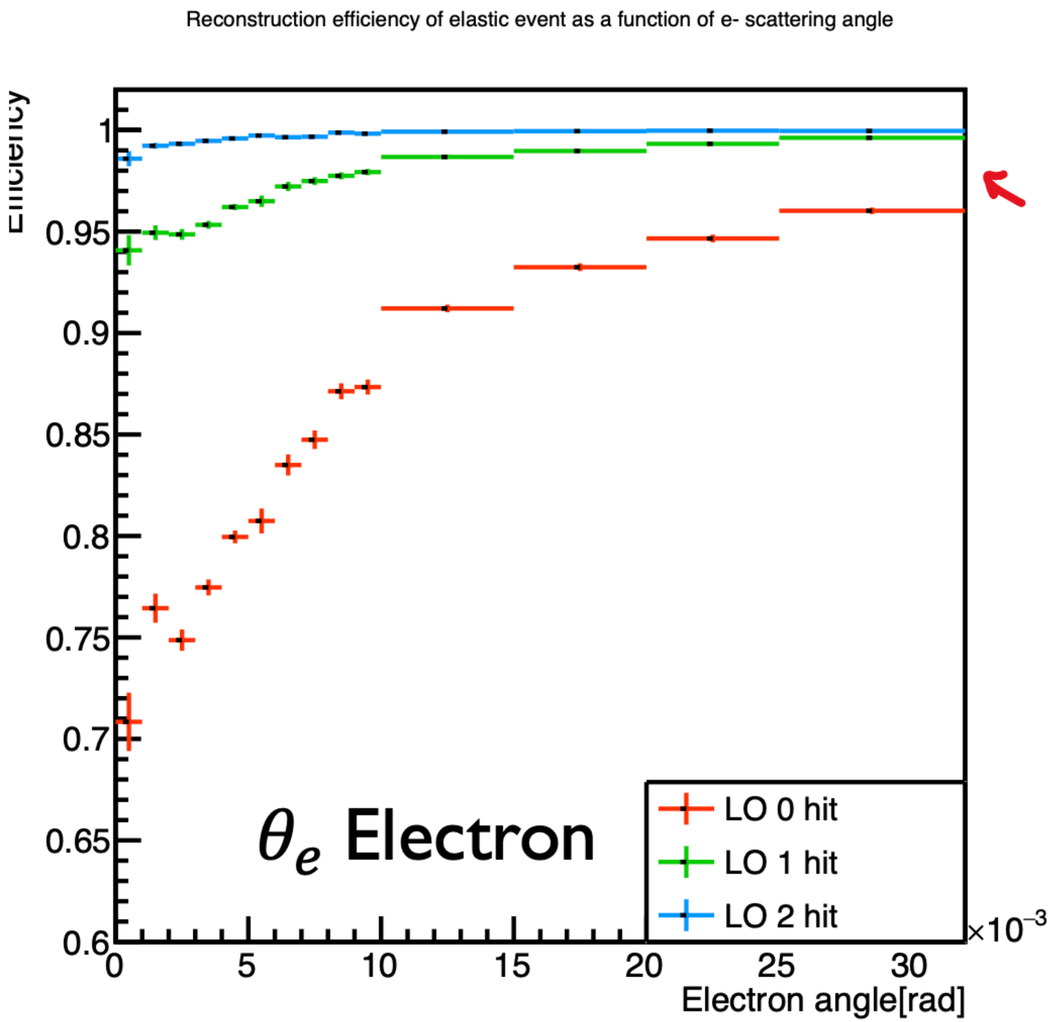




Plot generated using $QLT = 67\,\%$ and overflow and **`maxNumberOfSharedHits`** = 0

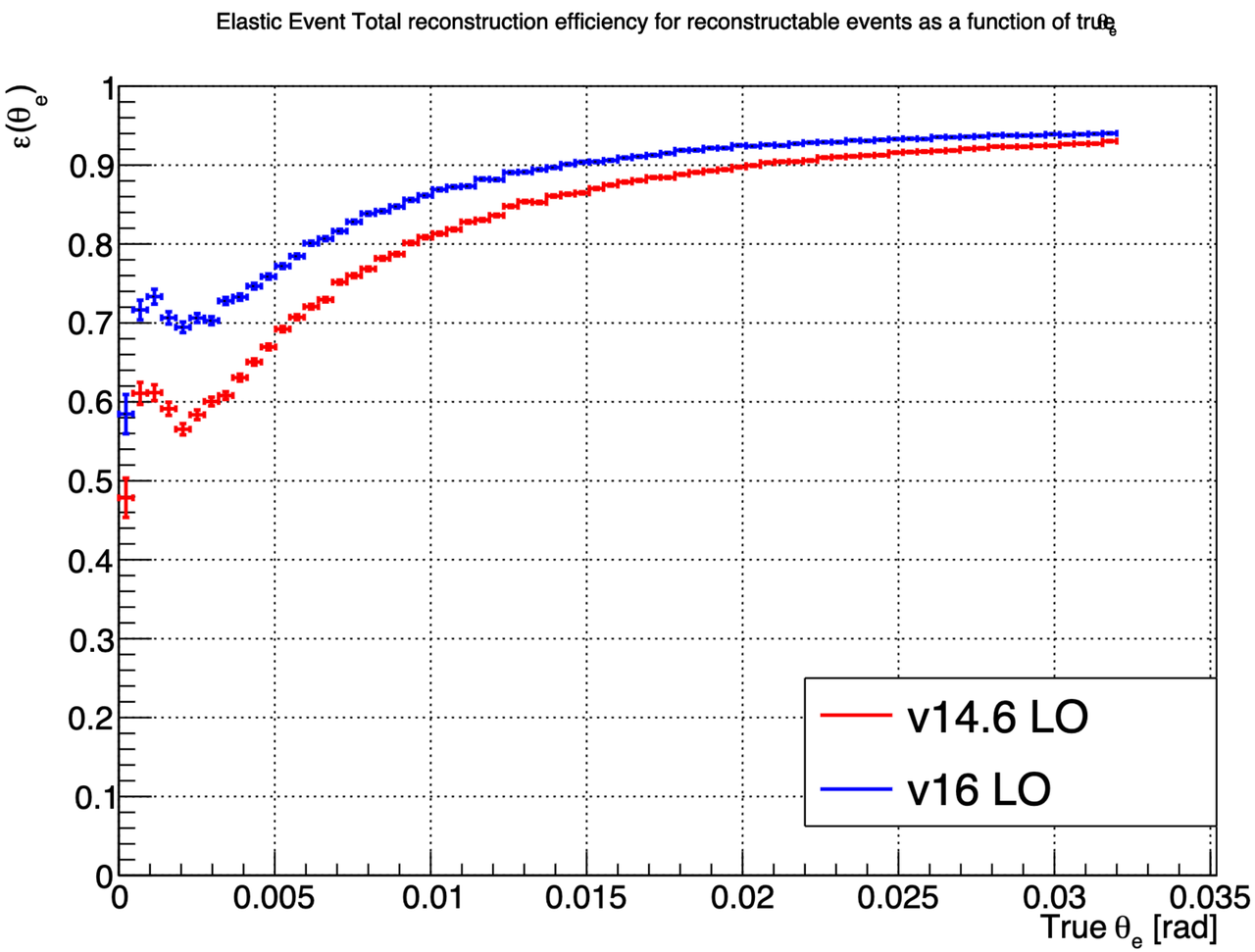
Eugenia's Reconstructable Efficiencies at LO

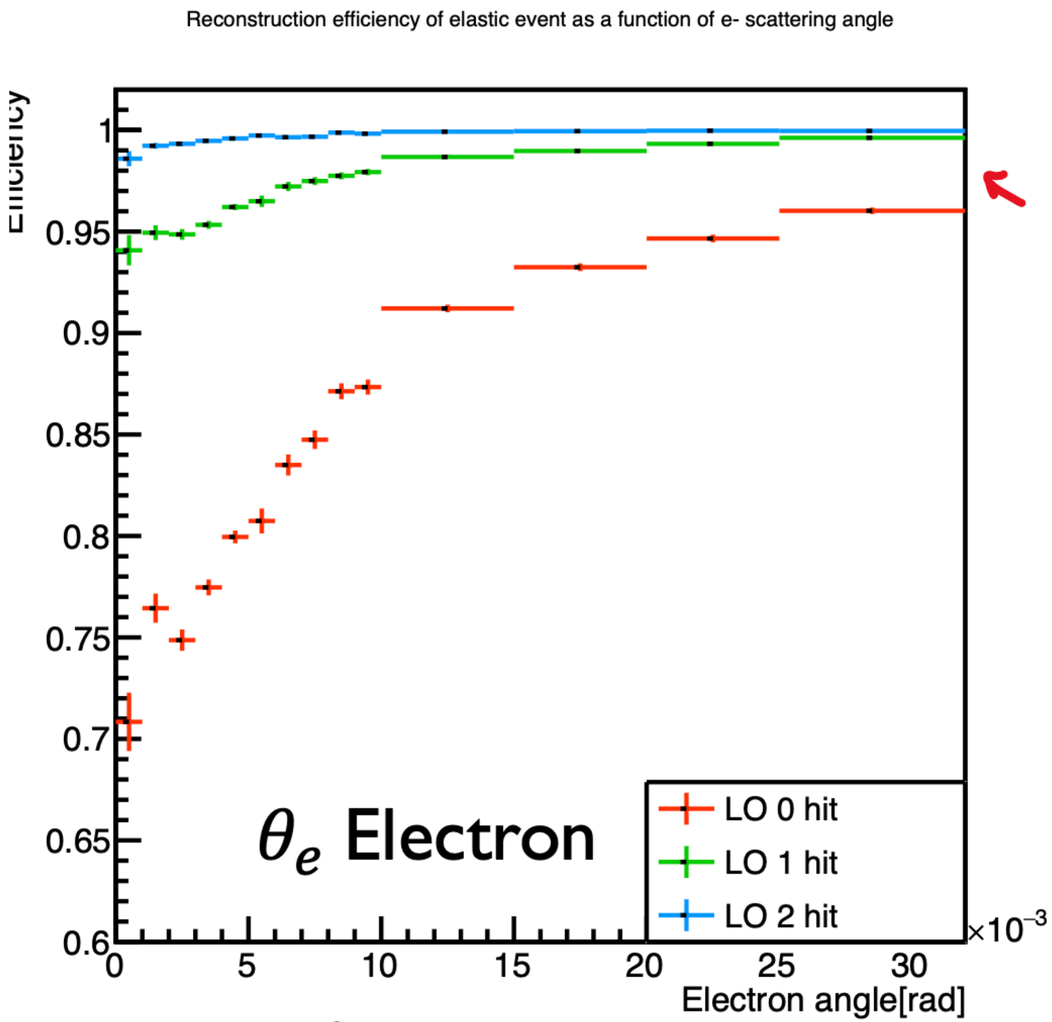




Plot generated using $QLT = 67\ \%$

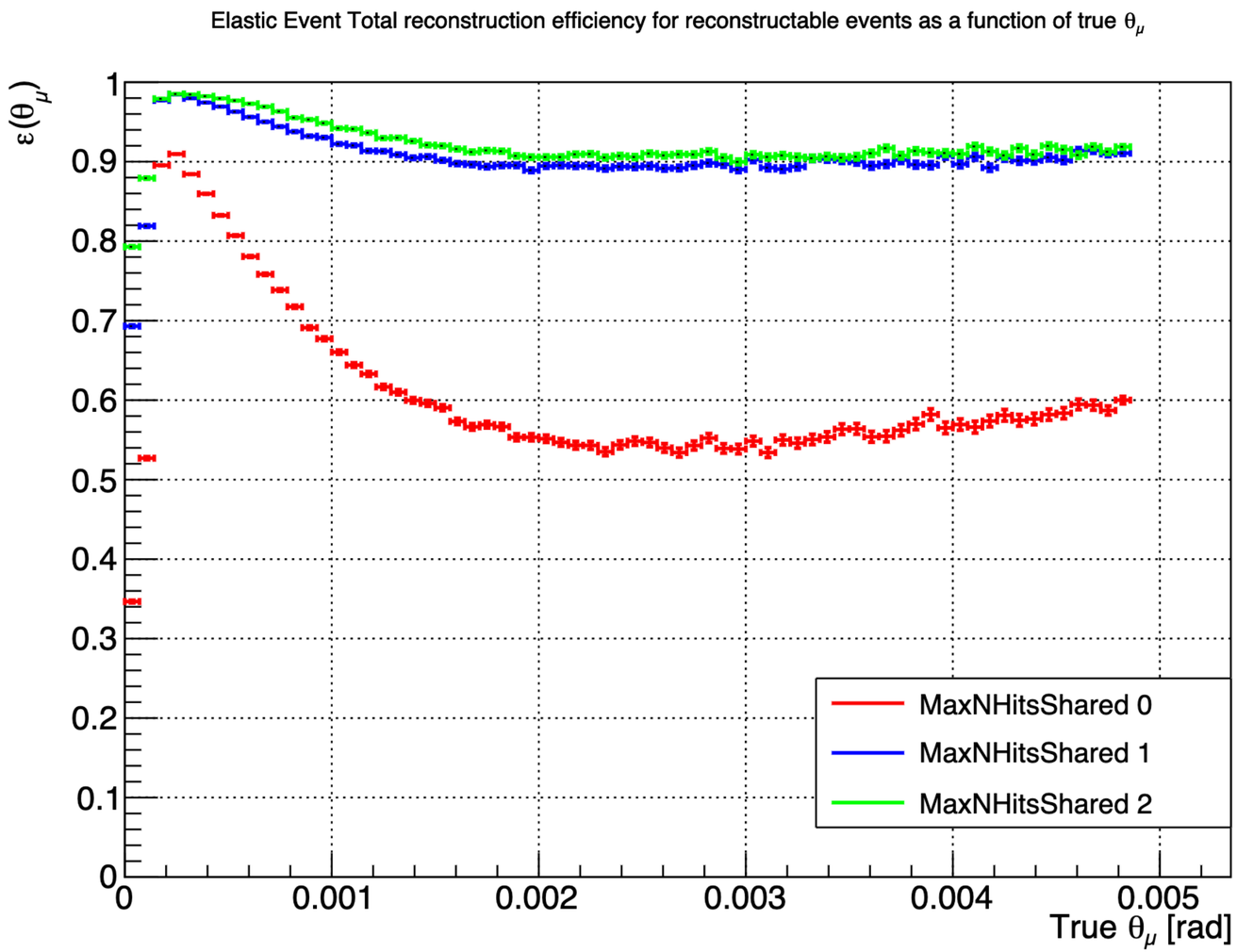
Eugenia's Reconstructable Efficiencies at LO



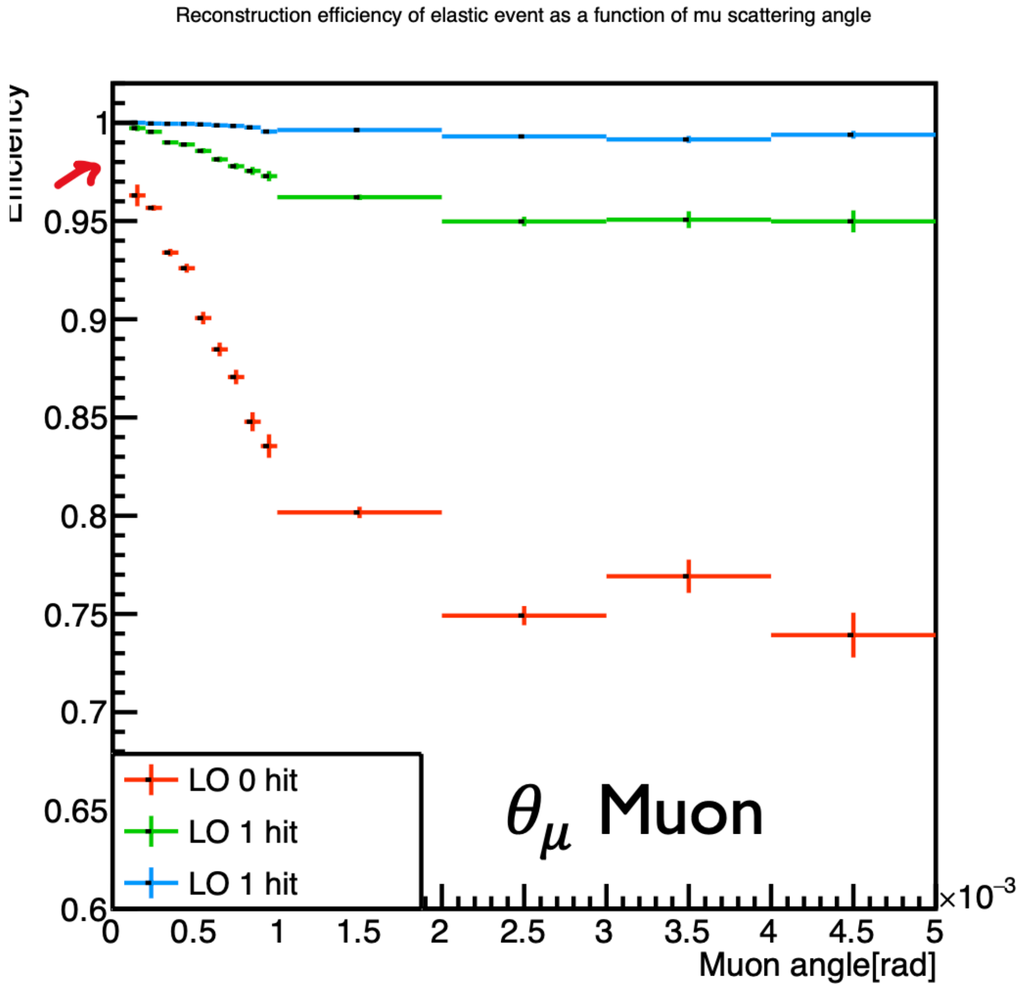


Plot generated using $QLT = 67\,\%$ and **`maxNumberOfSharedHits`** = 0

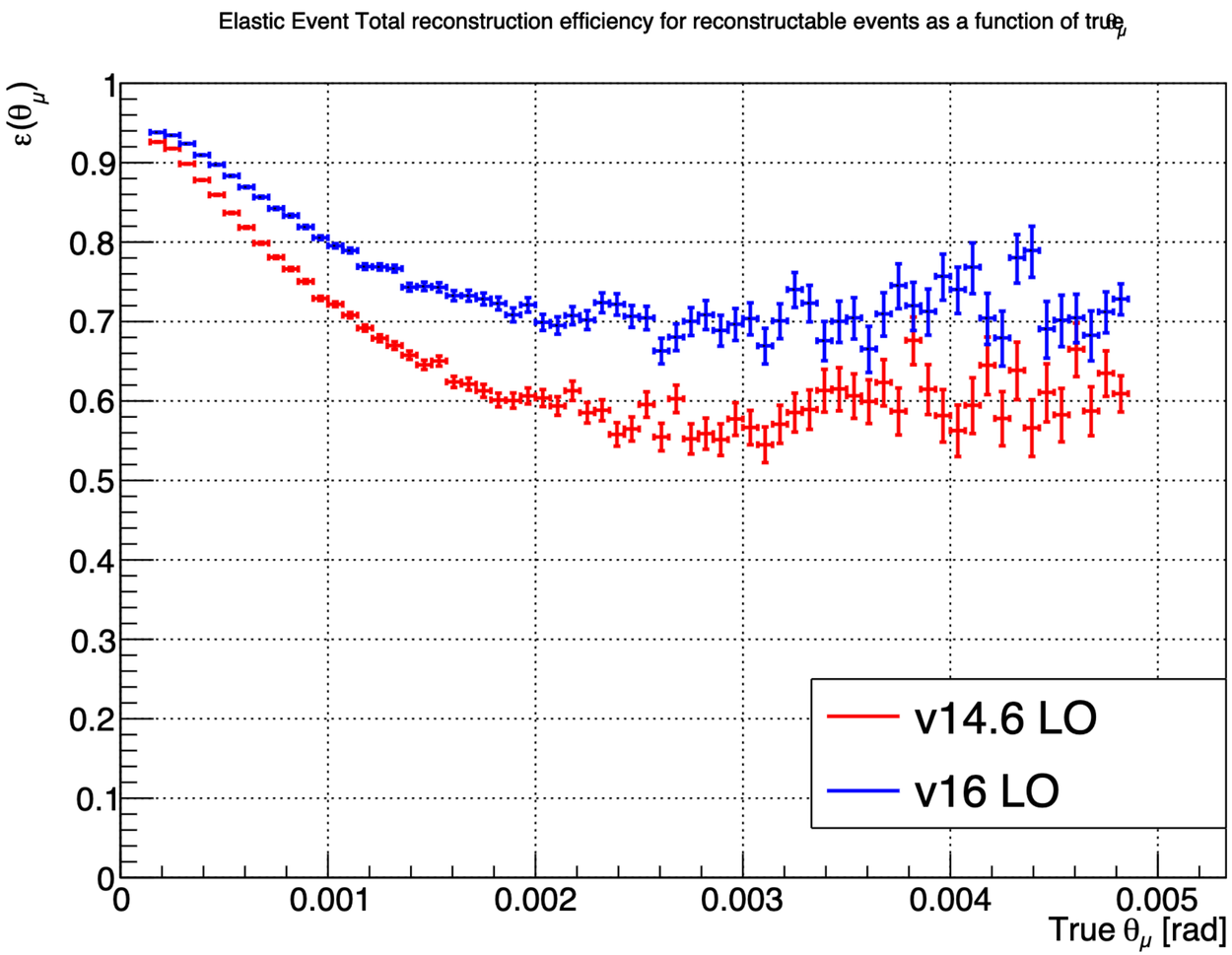
Eugenia's Reconstructable Efficiencies at LO

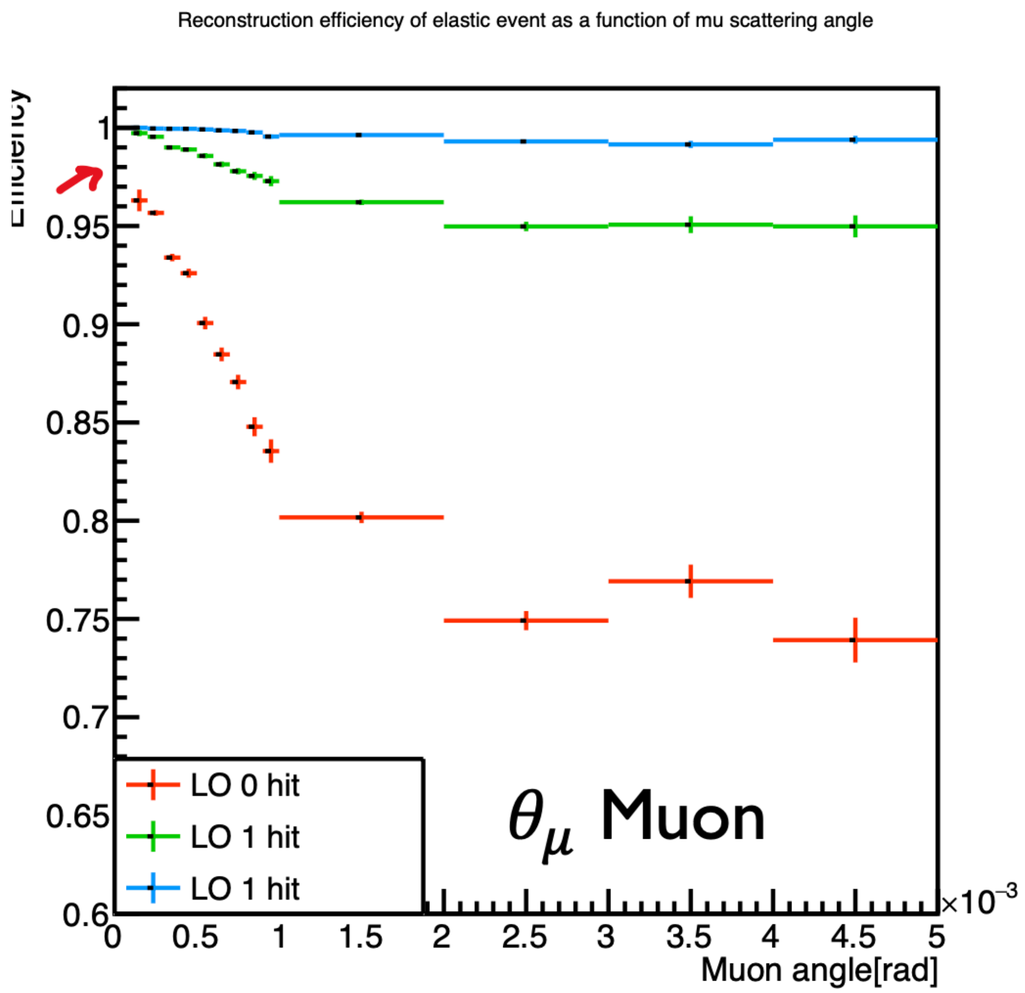


Plot generated using $QLT = 67\ \%$



Eugenia's Reconstructable Efficiencies at LO





Plot generated using $QLT = 67\ \%$

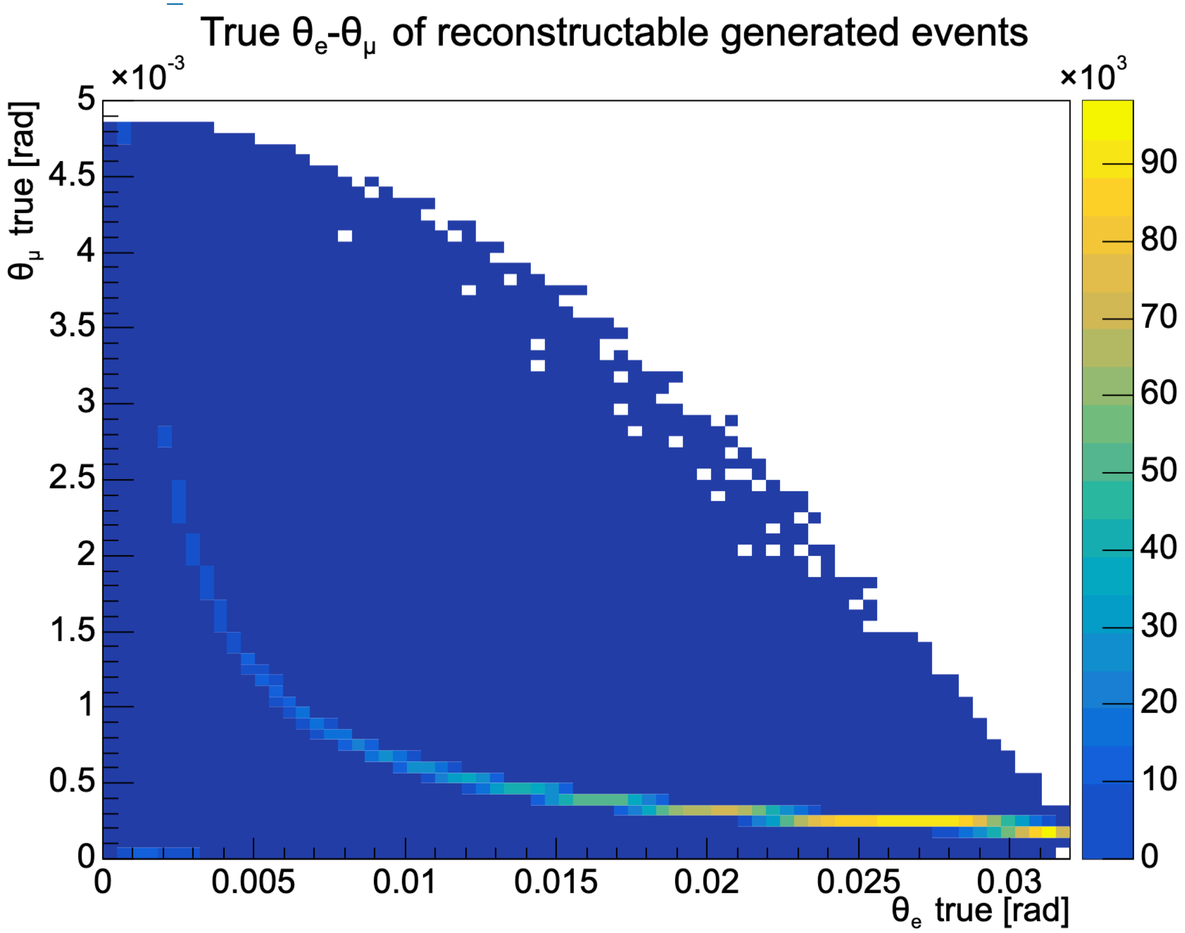
Conclusions

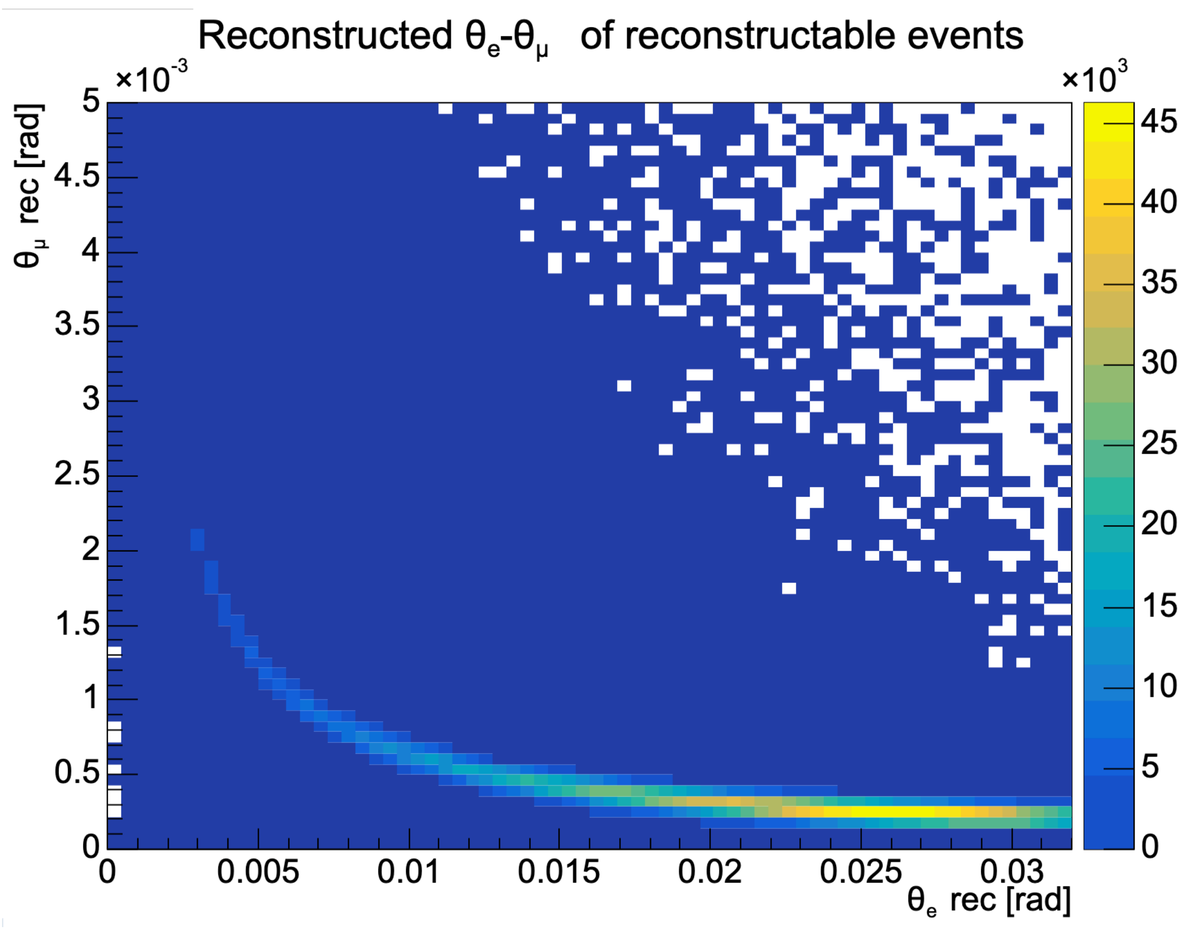
- I was able to replicate Eugenia' results at LO
- I expanded upon this by:
  - ✳ Going to NNLO → important drop in efficiency at low θ_e
  - ✳ Building smearing matrices to understand better what is happening
  - ✳ Simulating efficiencies of the simultaneous 2D (θ_e, θ_μ) distribution
- Next studies will be on the new 2025 configuration, with 3 tracking stations→ potentially track outgoing track with *two* stations

Thank You!

Backup

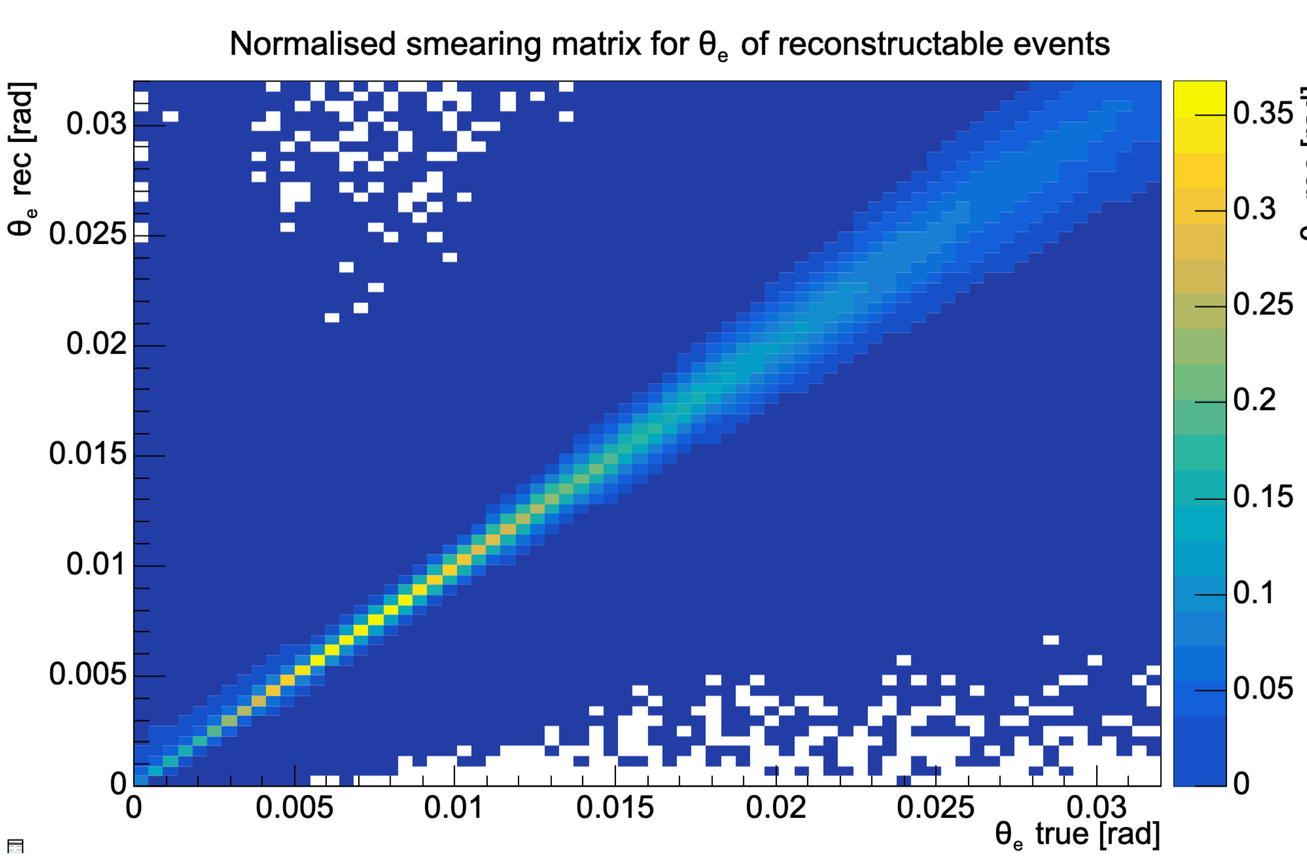
Generated and Reconstructed (θ_e, θ_μ) Distributions for Reconstructable Events

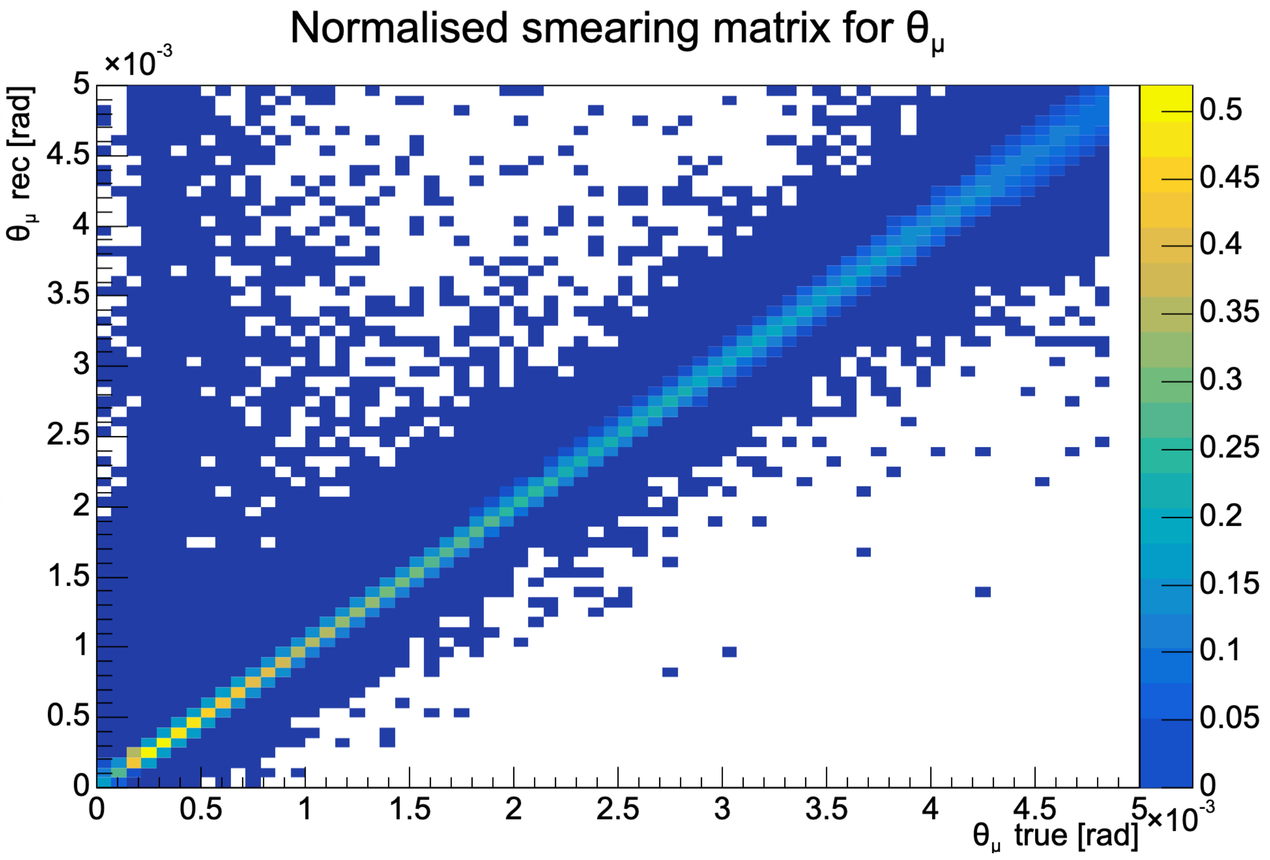




$QLT = 65\,\%$, MaxNsharedhits = 0, Reconstructable

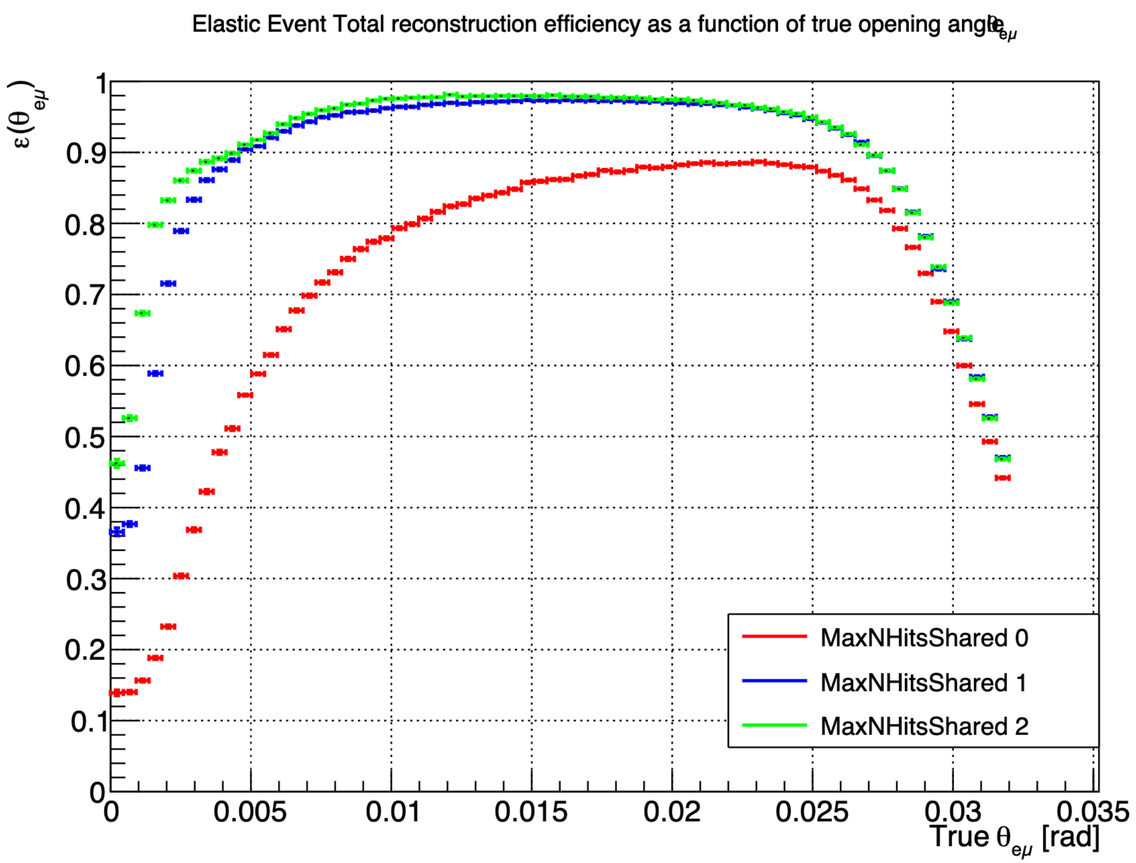
Normalised Smearing Matrices of θ_e and θ_μ for Reconstructable Events



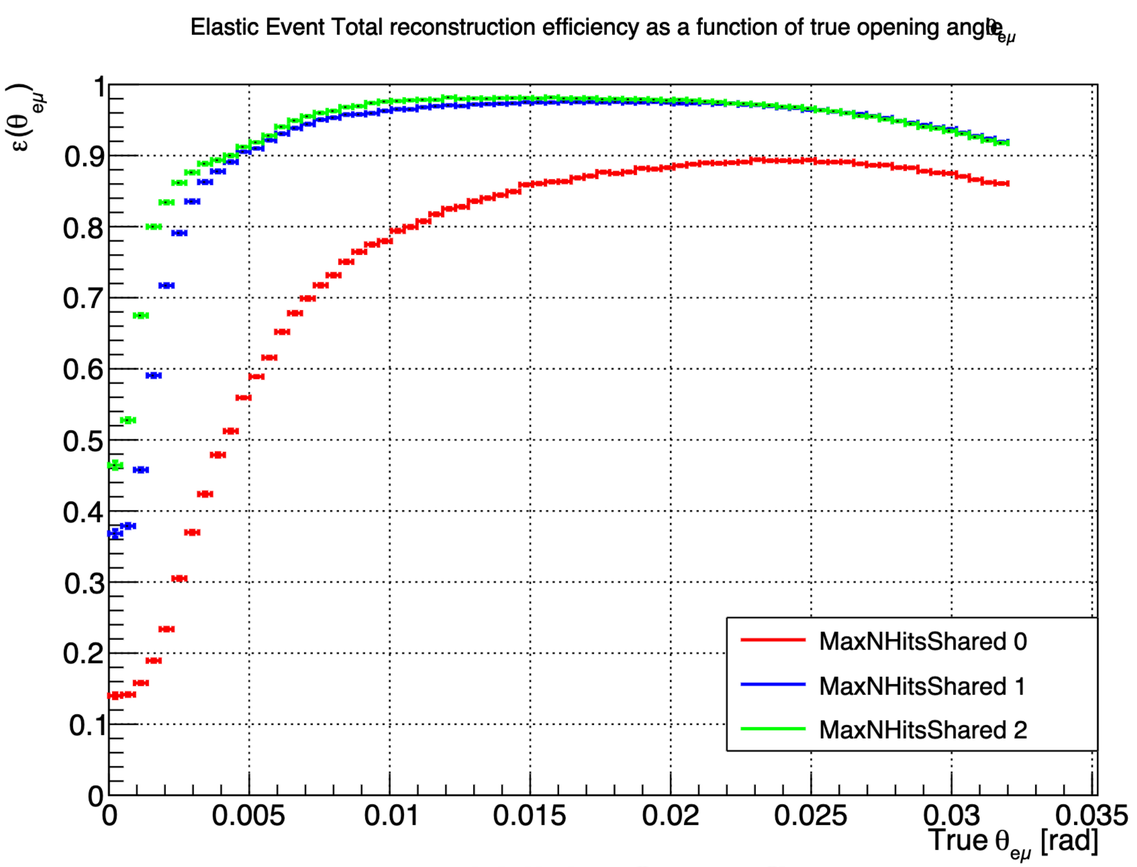


$QLT = 65\,\%$, MaxNsharedhits = 0, Reconstructable

Reconstructable Track Efficiency $\varepsilon_{trk,e}^{rca}$ as a function of $\theta_{e\mu}$



No overflow



With Overflow

Plots generated using $QLT = 67\ \%$

Use of `ReconstructedVertices` (Preliminary)