

# Status report on a different method of performing the ND Slicing, update

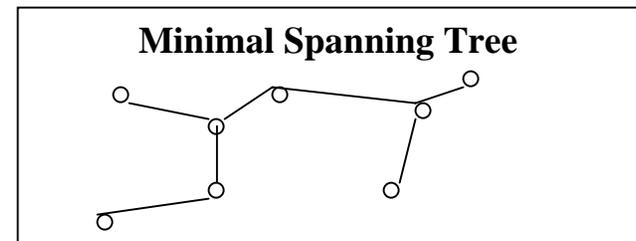
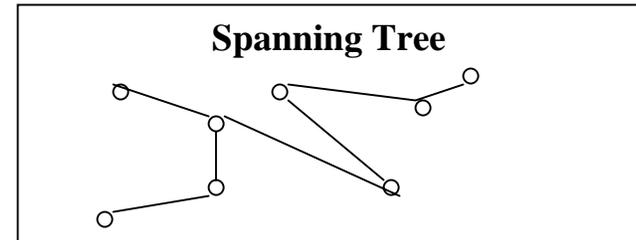
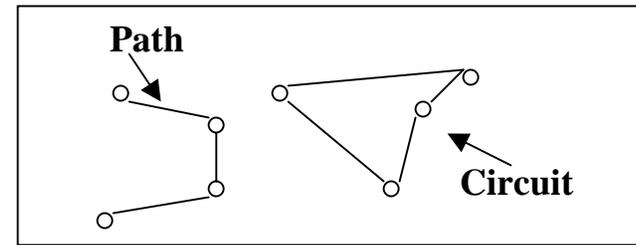
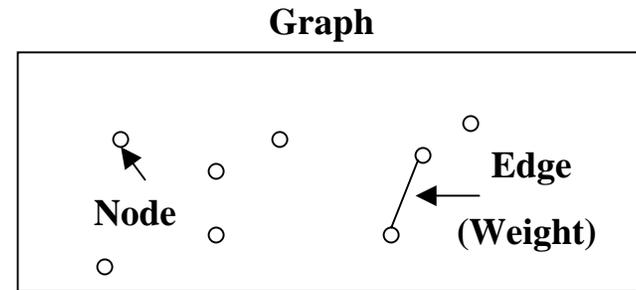
N. Saoulidou, Fermilab, 09-30-04

# Outline

- Description of the clustering technique (MST)
  - MST definition
  - Properties
  - MST & Clustering
  - Metric for weight definition
- Changes since Colab. Mtg. & comparison with current SR code :
  - Slice purity & completeness
  - Track purity & completeness
  - Shower purity & completeness
- Summary - On going work

# -Minimal Spanning Trees Basics-

- An **edge weighted Linear Graph** is composed of a set of points called **nodes** and a set of node pairs called **edges** with a number called **weight** assigned to each edge.



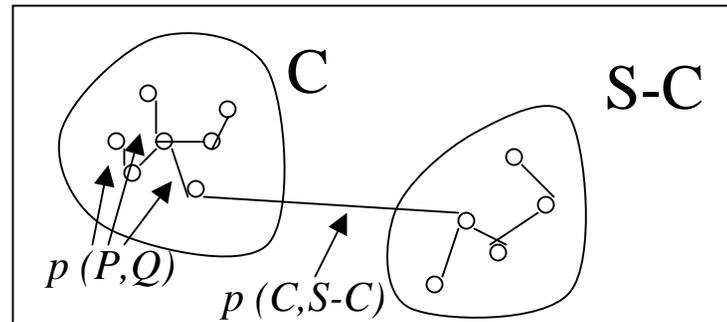
- A **path** in a graph is a sequence of edges joining two nodes. A **circuit** is a closed path.

- A **spanning tree** is a connected graph with no circuits which contains all nodes.

- A **minimal spanning tree** is the spanning tree whose weight (= sum of weights of its constituent edges) is minimum among all spanning trees in this set of nodes.

## - MST Theorem 1 -

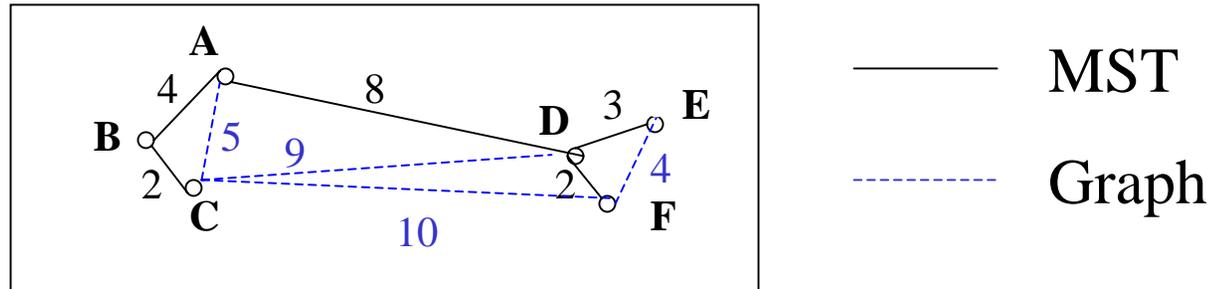
- Theorem 1 : " If  $S$  denotes the nodes of  $G$  and  $C$  is a nonempty subset of  $S$  with the property that  $p(P,Q) < p(C,S-C)$  for all partitions  $(P,Q)$  of  $C$  then the restriction of any MST to the nodes of  $C$  forms a connected subtree of the MST ". The significance of this theorem for cluster detection can be illustrated if the following figure which depicts the MST for a point set containing two clusters  $C$  and  $S-C$  :



- This theorem assures us that **the subgraph of an MST does not break up the real clusters in  $S$** , but on the other hand neither does it force breaks where real gaps exist in the geometry of the point set.
- A spanning tree is forced by its very nature to span all the points but at least **the MST jumps across the smaller gaps first**.

## - MST Theorem 2-

- Theorem 2 : "If  $T$  is an MST for graph  $G$  and  $X, Y$  are two nodes of  $G$ , then the unique path in  $T$  from  $X$  to  $Y$  is a minimax path from  $X$  to  $Y$ ".[1]



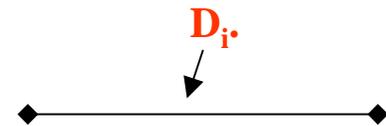
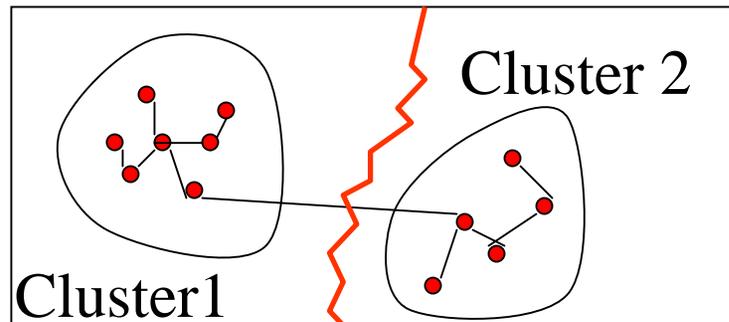
- **Cost** : maximum edge weight of the path e.g the path (CBADE) has a cost of 8.
- **Minimax path** : The path between a pair of nodes that has the least cost e.g there are four minimax paths from C to F all of cost 8.
- The **minimax path** each of whose **subpaths** are also **minimax** lies **within the MST** and that is not a coincidence as shown in the previous theorem.
- So the preference of minimax paths in the **MST** forces it to **connect** two **nodes X and Y** belonging to a **tight cluster** without straying outside the cluster.

## - MST properties -

- **The MST is deterministic.** It does not depend on random choices in the algorithm or on the order in which nodes and edges are selected and examined but only on the given set of nodes.
- **The MST is invariant under similarity transformations,** that is under all transformations that preserve the monotony of the metric (rotations, translations changes of the scale and even some nonlinear distortions).
- **The metric for the weight assignment** can be defined in many ways and does not have to be the **Euclidean Distance between 2 nodes**.

## -MST & Cluster Analysis-

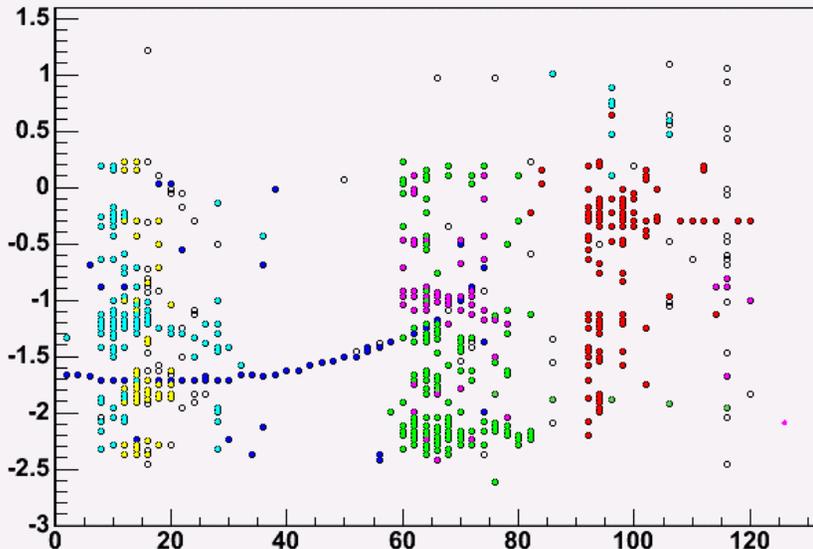
- Main Idea : *After forming the MST of a set of points group the points into disjoint sets by joining all edges of weights  $D_i$  or less. Each set is then said to form a cluster at level  $D_i$ . Thus all segments joining two clusters defined at level  $D_i$  will have lengths greater than  $D_i$ .*



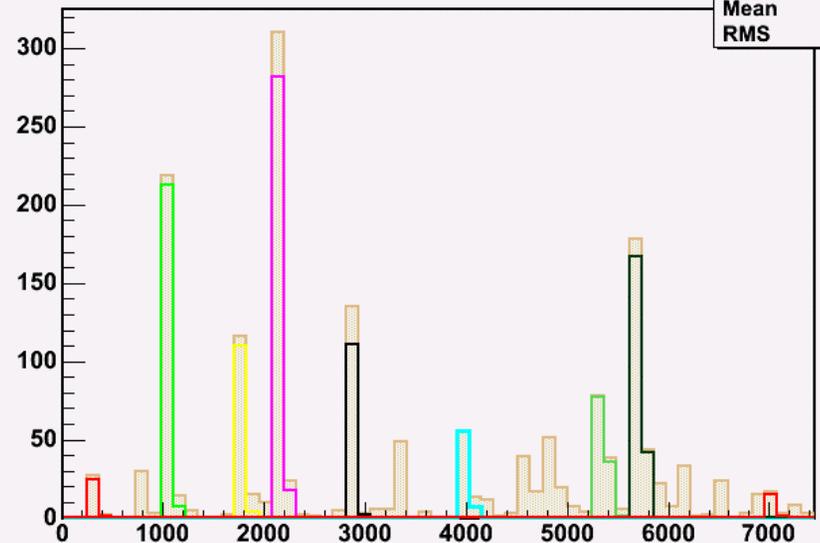
# MST in ND (Clustering in time & space)

- I have used a hybrid metric for constructing the length of the MST branch connecting two strips :
  - time difference  $\times c + a \times$  length difference (in z) and I using a semi-random initial cut that is going to be tuned...
  - This method has just ONE parameter that needs to be tuned...
  - Ok I lied ... it has 3 :
    - Minimum number of strips to form a slice
    - Minimum pulse height to consider the strip
    - Cut on the length of my TREE.

U (m) vs Z(m) Different Slices in Snarl 2



stp.time1\*10\*\*9 {fHeader.fSnarl==2}



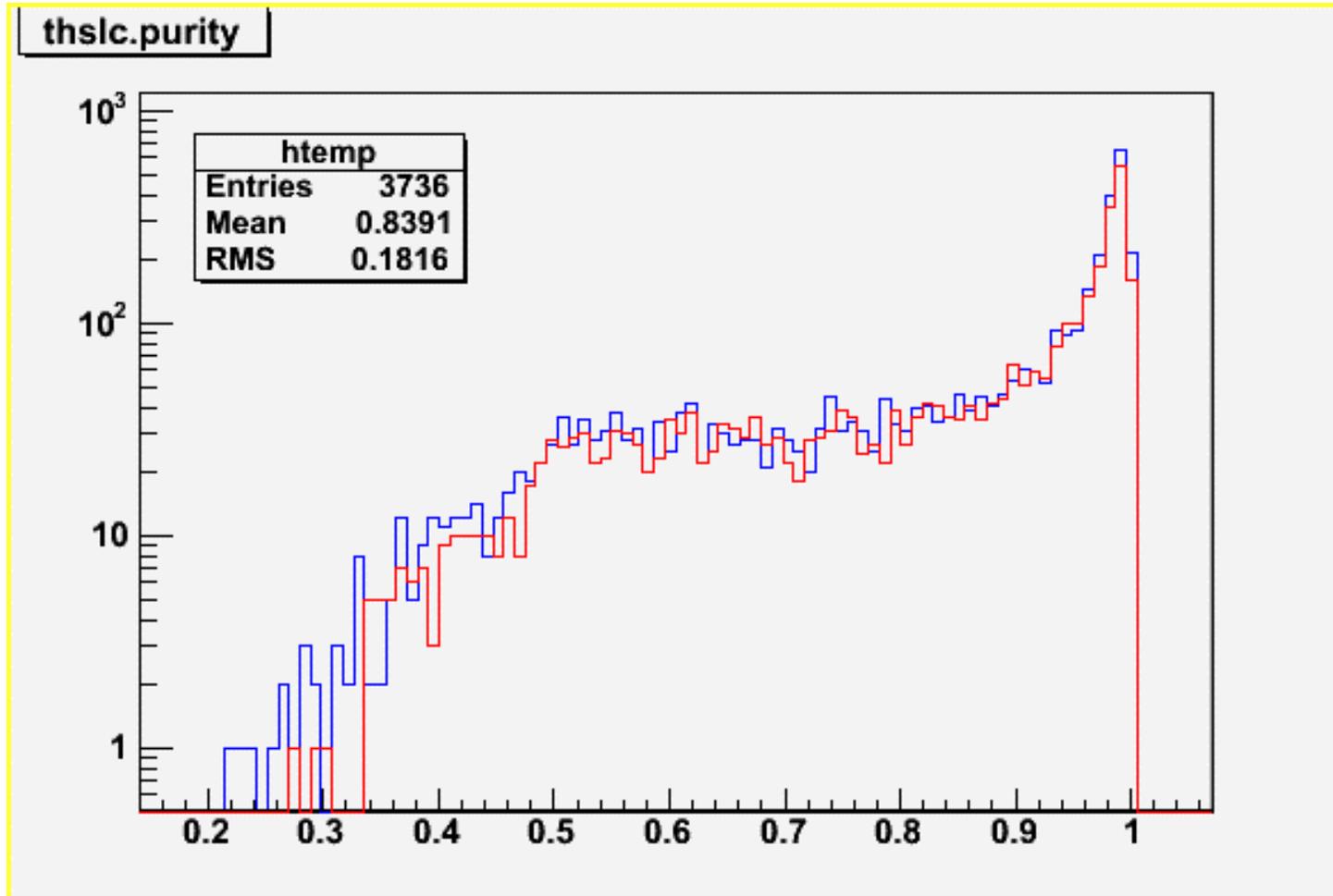
htemp	
Entries	1735
Mean	3371
RMS	1889

# Updates- changes since Colab. Mtg.

- Changed the minimum number of strips to form a slice (this is a global parameter of my code as well) from 10 to 5 after Tom O. discussed the implication of that the NC events
- Changed the Pulse height cut in order to consider a strip as member of a slice from 2 PEs to 0 PEs (I am not cutting anything).
- Changed the constant that Nathaniel uses in the MC to tune the light level from 1104 to 894 (!) as he suggested in the Colab. Mtg.
- Changed "a little" bit the way the time stamp of digits is used from the offline software for the ND (still experimenting)...
- Treat calorimeter and spectrometer strips in exactly the same way with my code.
- Run my code with all these changed and the SR code with the changed related with the "Minimum number of strips" and "Pulse height cut" changes and compare....
- **The comparison is ONLY based for the moment on truth helper quantities...I will do the same in much more detail if I figure out a way to "NICELY" align ND Tree entries, and run 200 events altogether without crashing (for the moment I run portions of 50 events that is faster and does not crash but creates trees that I have to chain after and the AddFriendByKey does not work.**

# MST results & comparison with SR

Slice purity MST (BLUE) SR (RED)

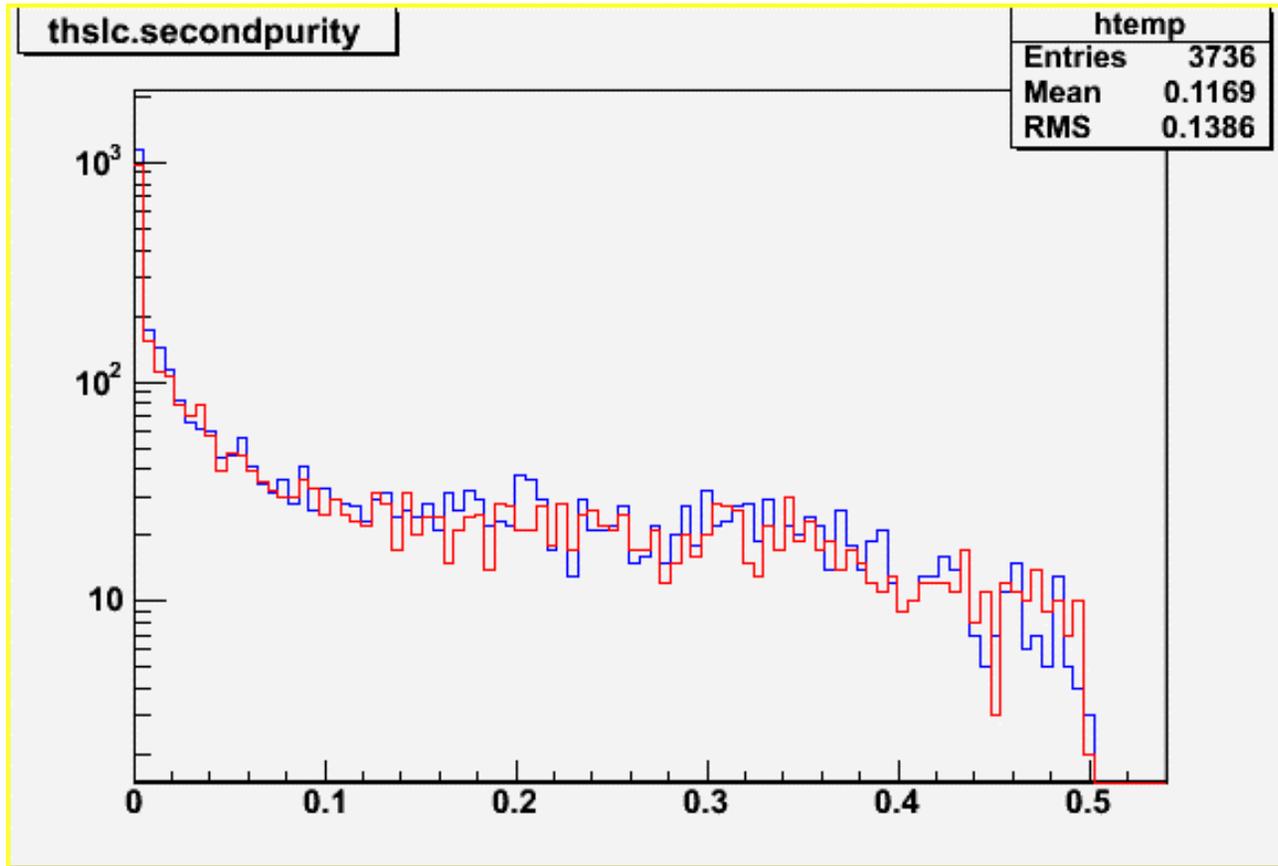


- The MST is as good as the SR(or better) ... MST finds more slices...(most probably most of them in the Spectrometer region)

# MST results & comparison with SR con't

Slice contamination from a second neutrino event

MST (BLUE) SR (RED)

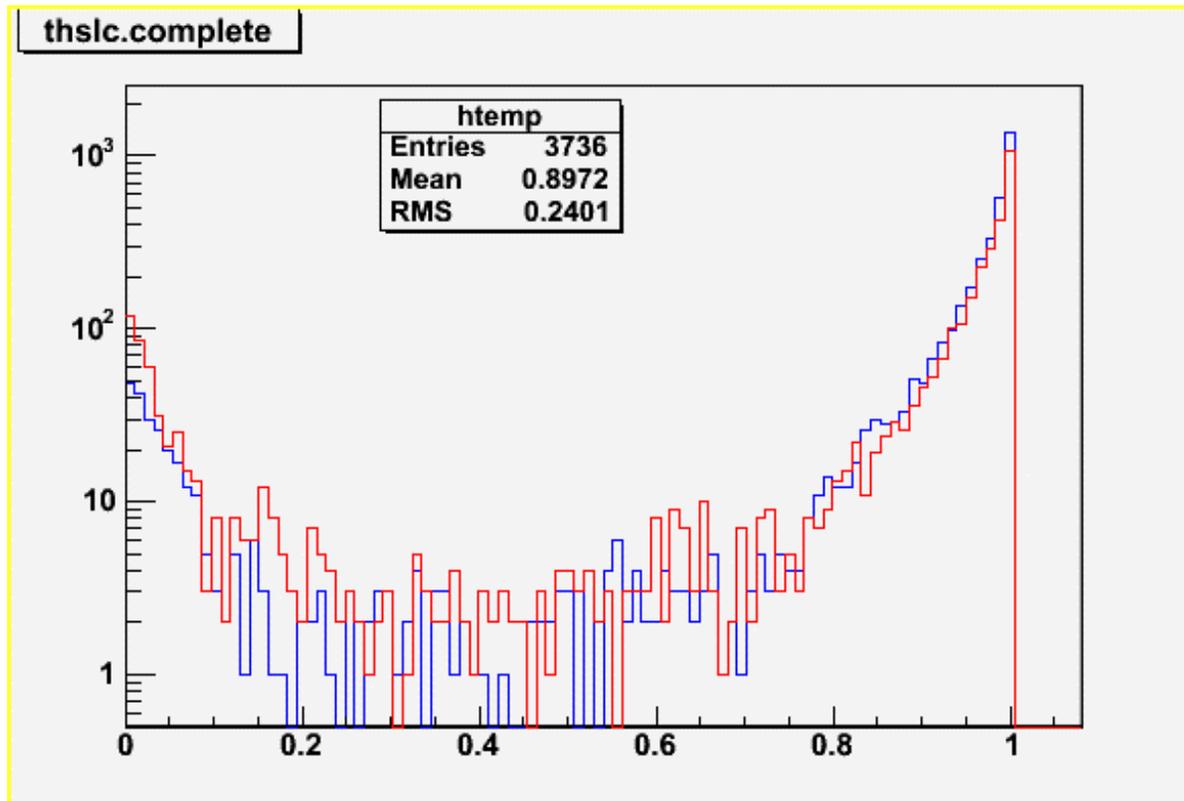


- The MST is as good as the SR...(or maybe better)

# MST results & comparison with SR con't

## Slice Completeness

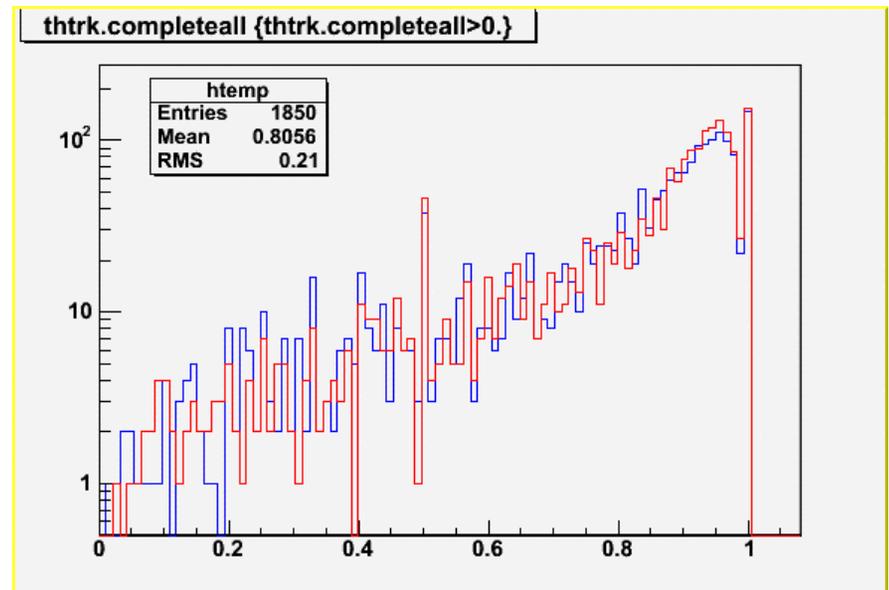
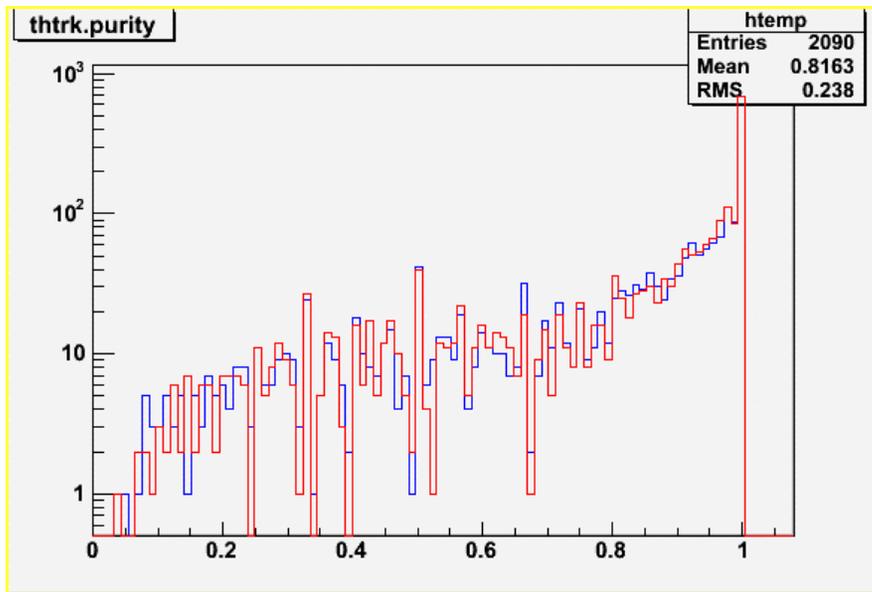
MST (BLUE) SR (RED)



- The MST is as good as the SR...(or maybe better)

# MST results & comparison with SR : tracking

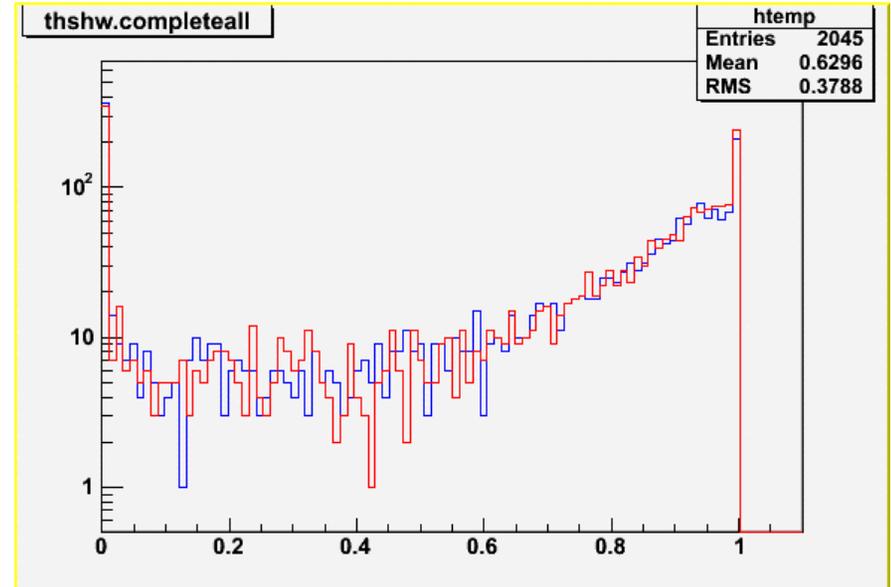
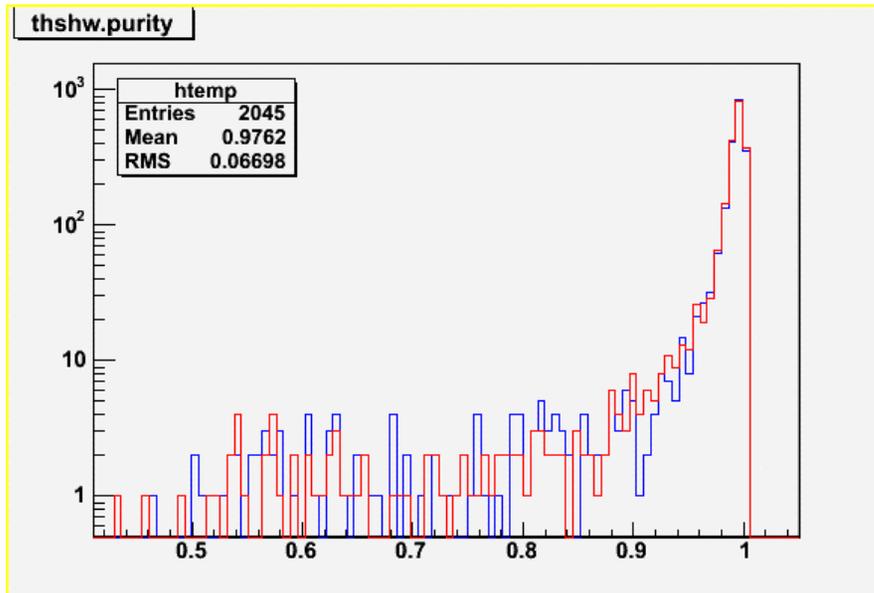
## Track purity & completeness MST (BLUE) SR (RED)



- The MST is as good as the SR ...

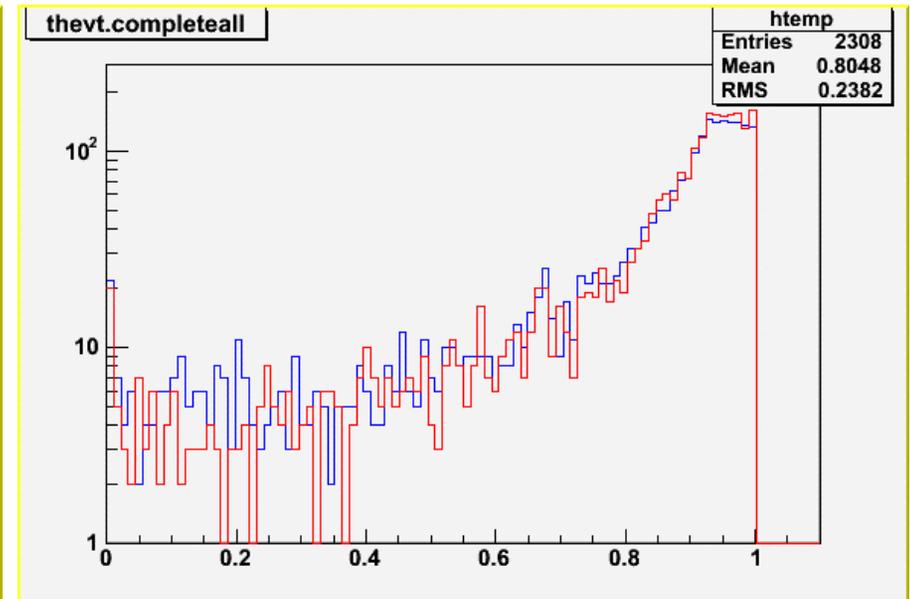
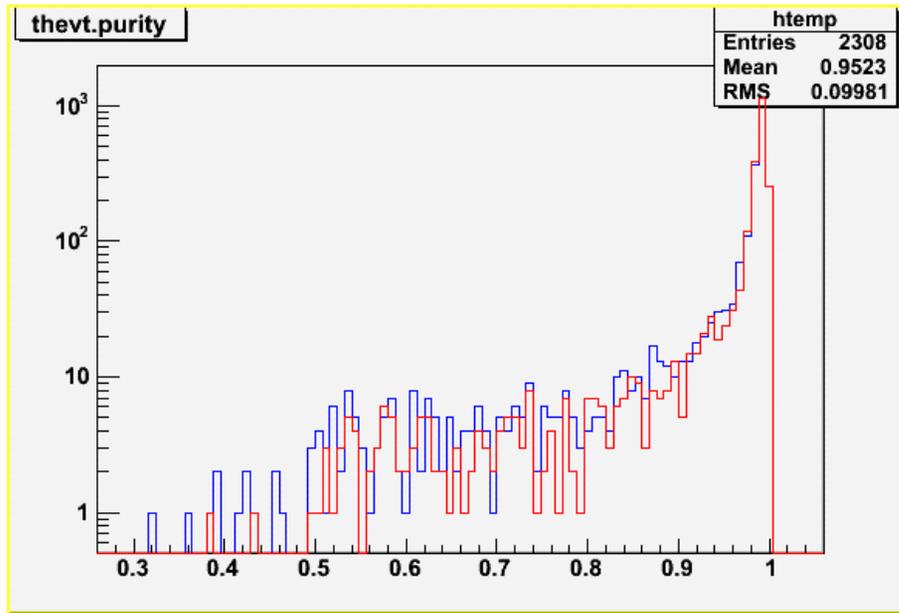
# MST results & comparison with SR : showers

## Shower purity & completeness MST (BLUE) SR (RED)



- The MST as good as the SR...

## Event purity & completeness MST(Blue) SR(RED)



- The MST is as good as the SR....

# Summary & Ongoing work

- This new slicing method is very promising and simple (results very very close to SR)
- It is flexible due to the freedom one has on the metric definition for the weight assignment
- It can be used to either strips or digits...
- It is ready to be used now...(I will send my code to Jim and then we can discuss it).
- I want to "play" a little bit more with ND timing as far as strip formation is concerned and also with my metric and cut...
- Next time I plan to have more detailed results on the various reconstructed quantities and on event selection efficiencies and biases (CC/NC)