NC Event selection (ND & FD) & Improvement of NC shower energy resolution

N. Saoulidou, Fermilab, NC Phone Mtg 02-05-05

Outline

- **NC / CC Separation (Update & slight improvement)**
  - Method (ANN)
  - Results
  - Statistical & systematic errors (work in progress)
- **NC Shower energy resolution using ANN (Update & slight improvement)**
- **Summary and ongoing work**

Cuts & MC events used

- Used all available MC R1.12 files. (using Christmas processing only for FAR, NEAR update in the next meeting)

- Fiducial cuts for ANN results :
  - **ND :1m around the beam center and 0.4 < mcz < 6.5**
  - **FD :3m around the detector center and z > 1 && z < 14 OR z> 17 && z < 29**

Shower energy estimation using ANNs (NC events)

- Started working on estimating shower energy using Neural Networks for Near & Far.

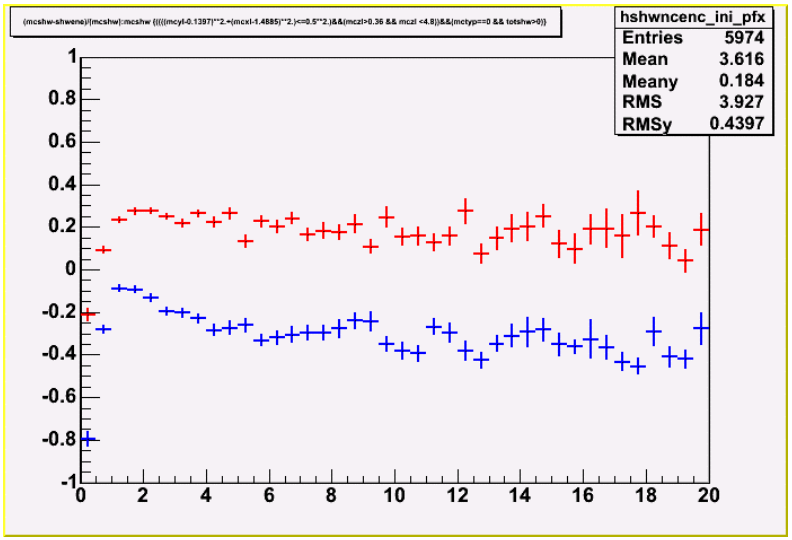
- Did a separate estimation for shower energies for CC and NC events. Ideally CC and NC showers should be the same. In reality tracks that are found in CC events do not share hits (strips) with the formed showers which results in an underestimating shower energies in CC events more frequently than in NC.

- Advantages:
  - Better energy resolution.
  - Better agreement between Near and Far detector (NNs "learn" the reconstruction differences and how to incorporate those in the final "answer").

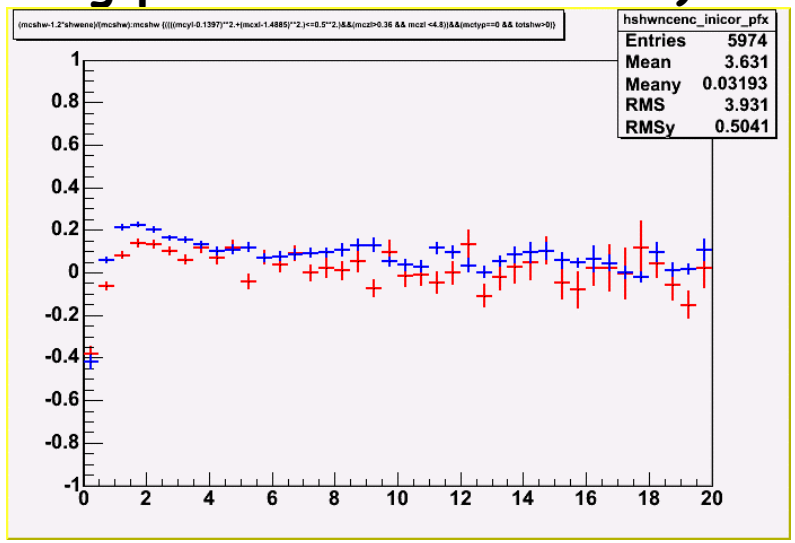
- Disadvantages:
  - General NN disadvantage : MC must describe data well.
  - Current disadvantage : Trained with actual energy distribution, flat energy distribution would be much better in our case since low energy events, that are of great interest are very very few and the NN learns better the high energy ones that are the majority.

Shower energy : initial results (code) & correction

dE/E vs Eshower (Old results using pre Christmas MC files)



Initial

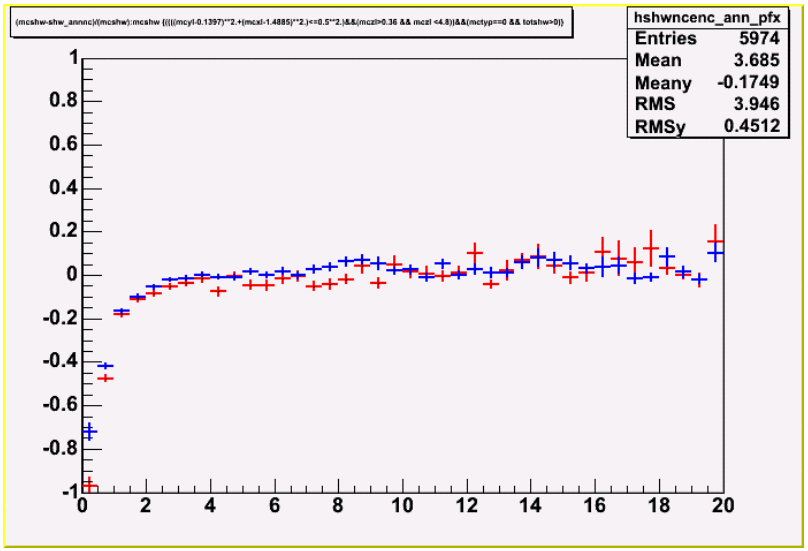


Corrected

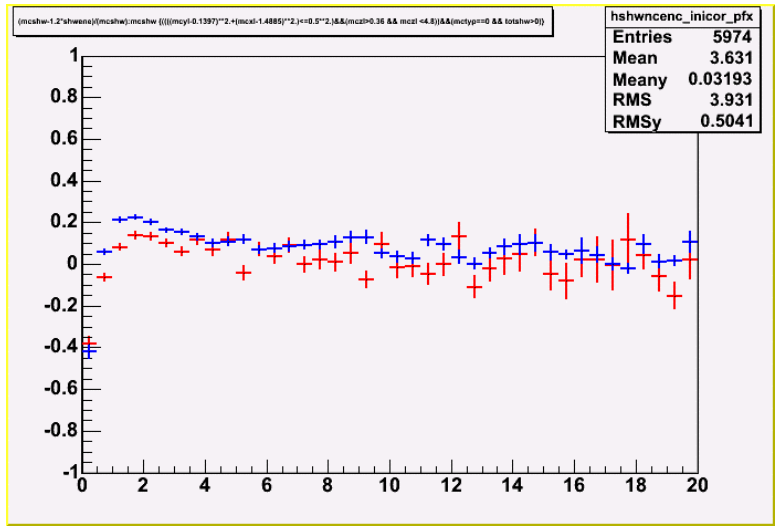
COLOR CODE : RED NEAR - BLUE FAR

- The initial DE/E for Near and Far showed a different offset. Corrected introducing a different multiplication factor.
- The initial DE/E shows a large overestimate for low shower energies.

Shower energy : NN results dE/E vs Eshower



NN estimation



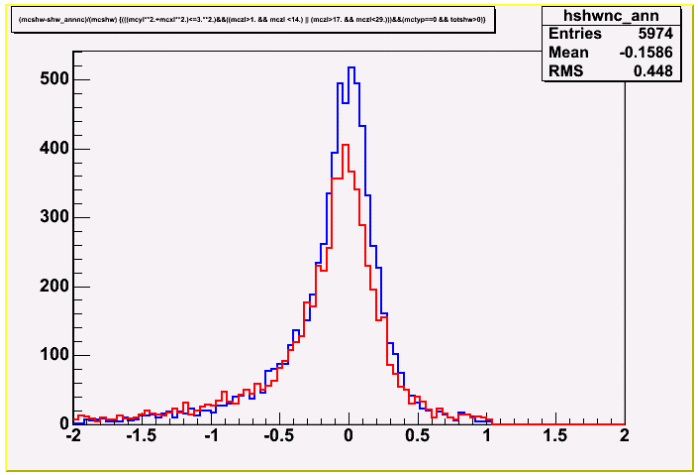
Initial corrected

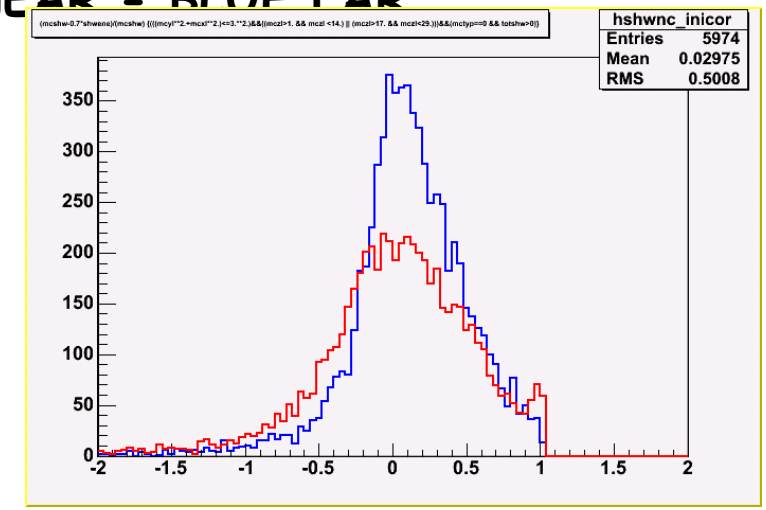
COLOR CODE : RED NEAR - BLUE FAR

- The NN results (which are quite preliminary) show :
  - Much better agreement between Near and Far
  - Much better energy resolution
  - Overestimate of shower energy for low energies.
- A flat energy distribution might give better results and also a better track-shower energy hit reconstruction and a more "clever" NN variable selection.

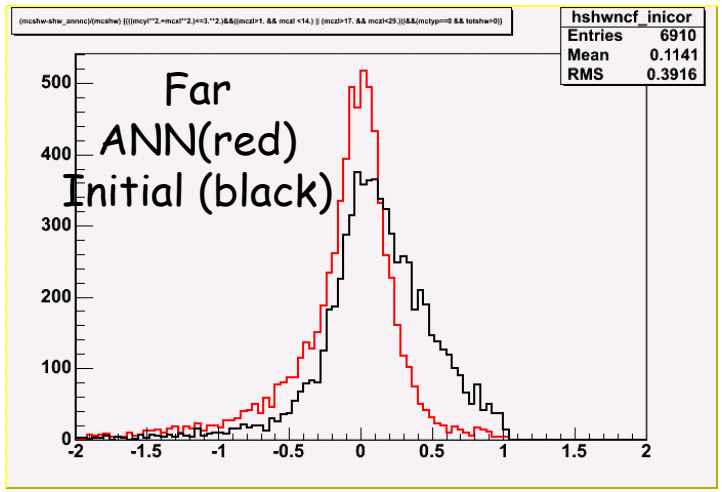
Shower energy : NN results Energy resolution (dE/E)

COLOR CODE : RED NEAR - BLUE FAR

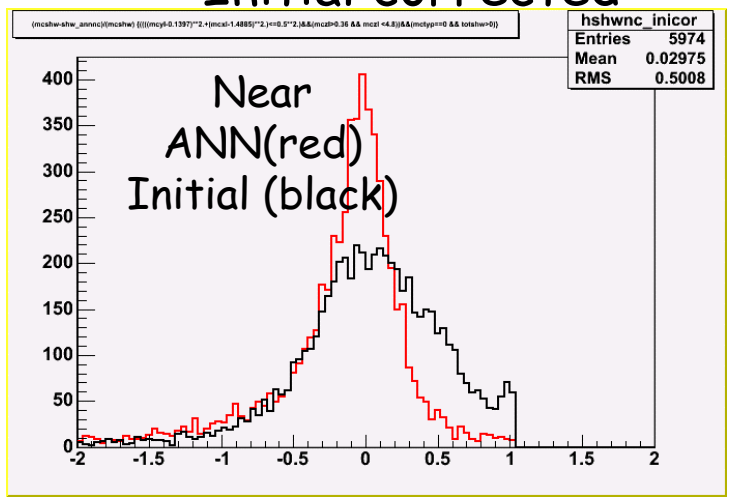




NN estimation



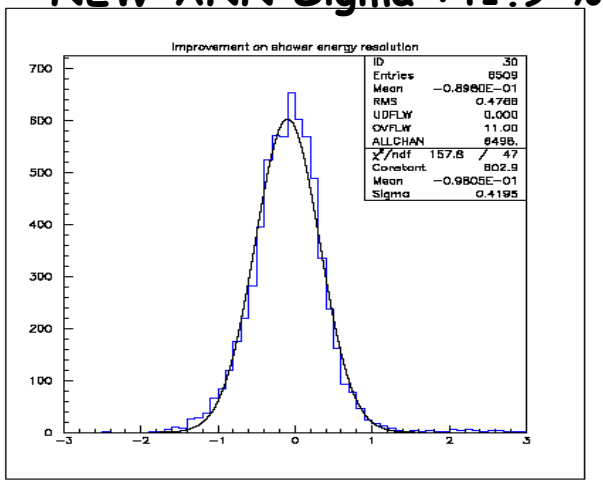
Initial corrected



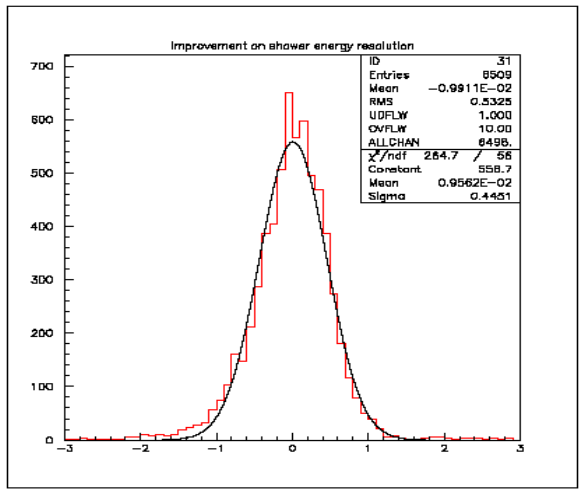
- The NN improve the energy resolution for both Near and Far and also makes energy estimation (relative calibration) better between Near and Far.

Shower energy : NN results Energy resolution (dE/sqrt(E))
FAR NEW "Christmas" MC files

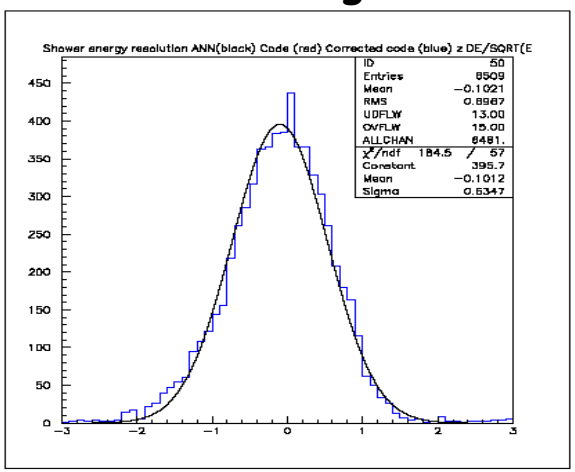
NEW ANN Sigma :41.9 %



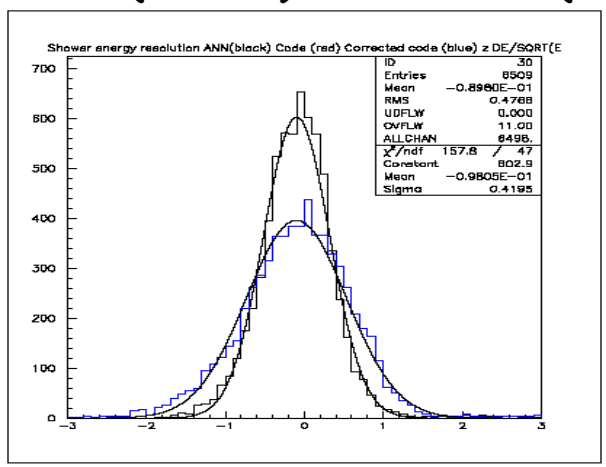
OLD ANN Sigma :44.5 %



NEW "CODE" Sigma :63.5 %



CODE (BLACK) NEW ANN (BLUE)



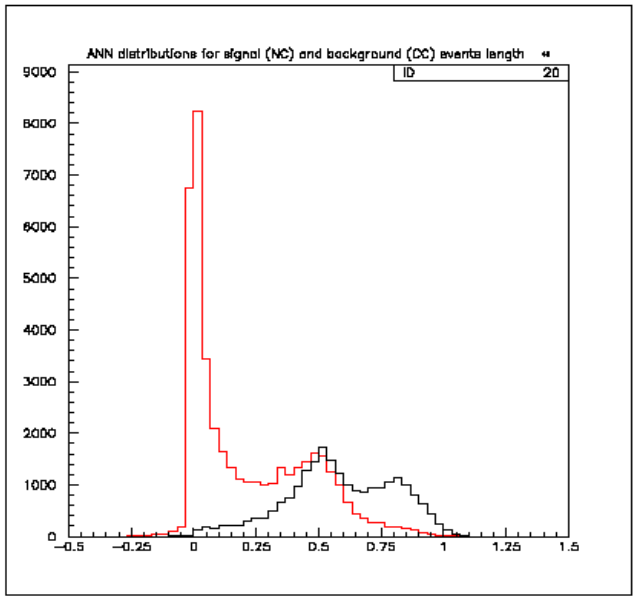
- ANN NC shower energy resolution further improved. Still more work to do…ANN results clearly much better than offline Code…

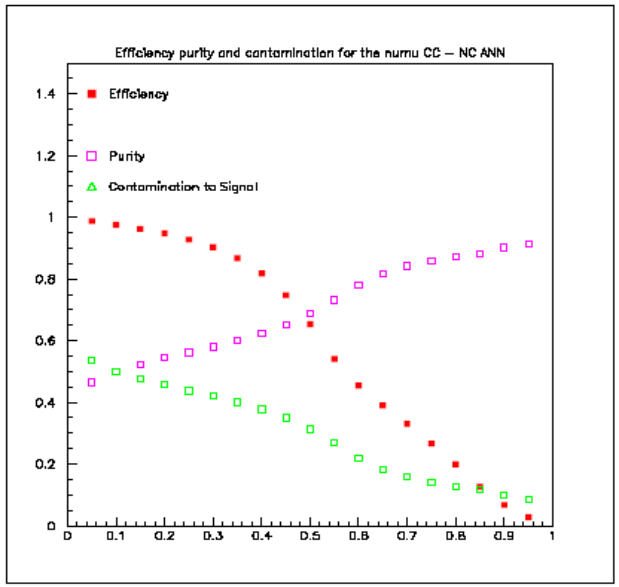
Event Selection

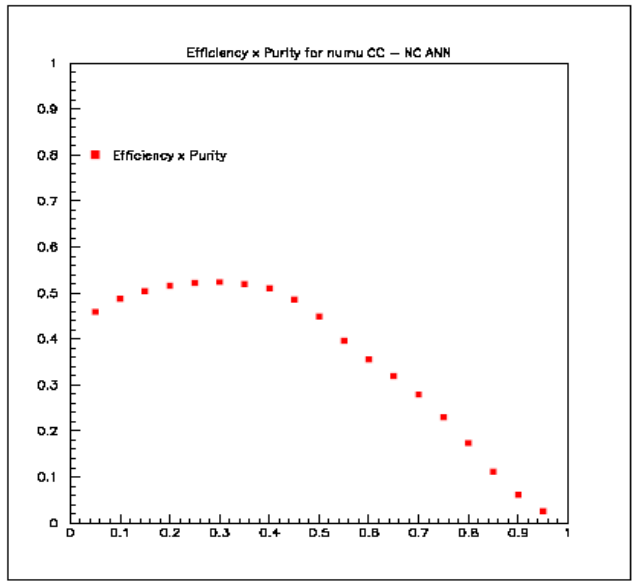
- Event selection method : ANN
- Train ANNs in :
  - NEAR
  - FAR unoscillated
  - FAR oscillated with 2 different dm2 (0.002eV^2 and 0.0025 eV^2 with sin^2(2theta) = 0.95
- Training with different "oscillation" scenarios was performed in order to study the effect of oscillations to the classification procedure and estimated the systematic uncertainties from that.

ANN Results NEAR (length<40)

A priori probabilities 1:1.9



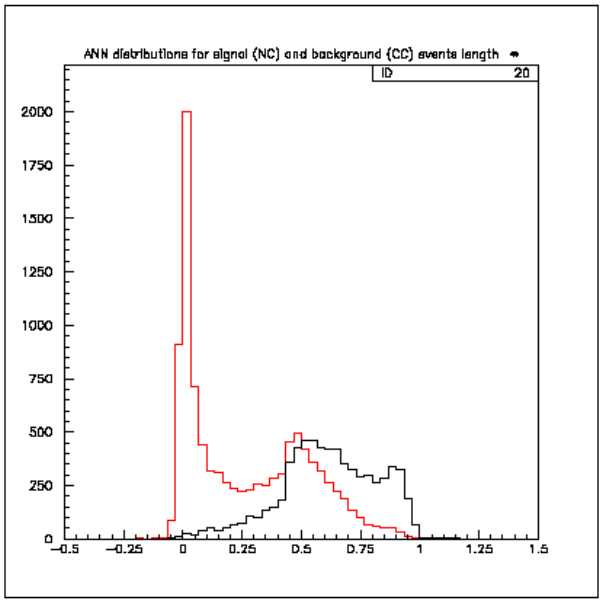


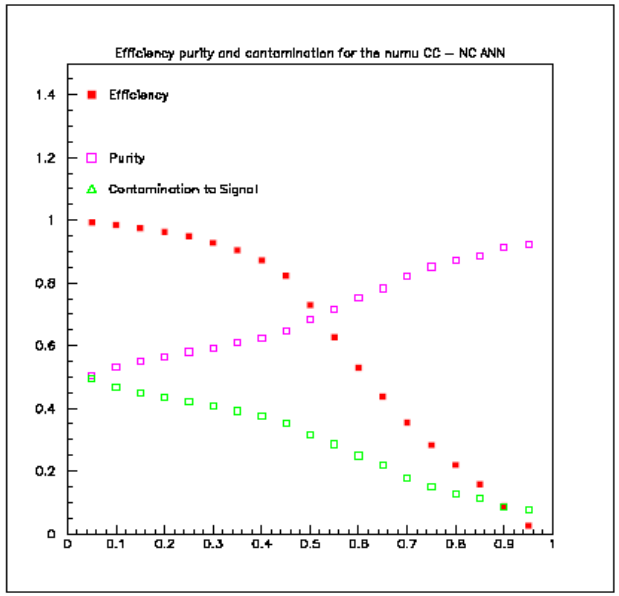


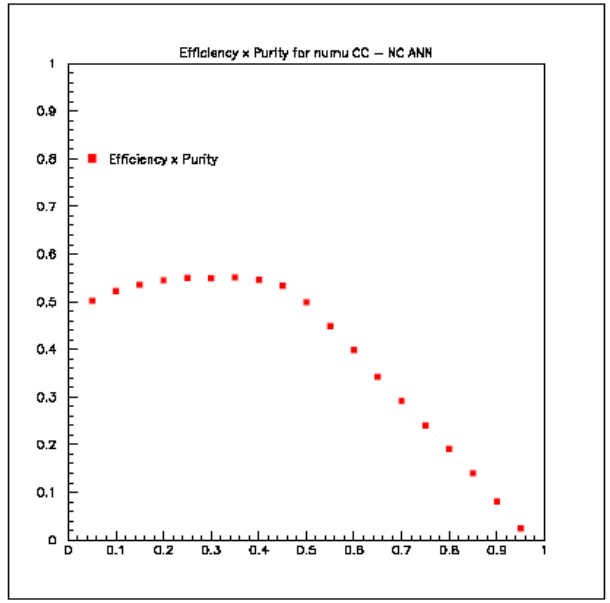
Event Probability Efficiency & purity for NC events Efficiency x Purity

- The ANN performs as expected : Higher purity for CC selection and lower for NC selection.
- If we set the cut @ 0.45 (i.e) we have an efficiency of 75% and a purity of 65%.

ANN Results FAR no oscillations (length<40)





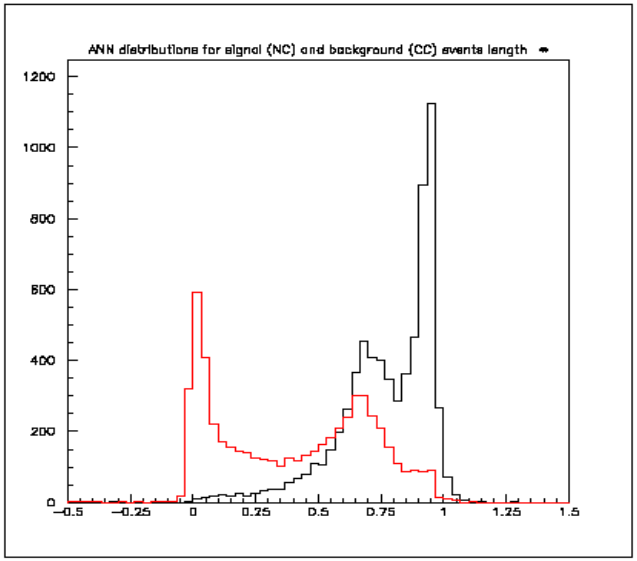


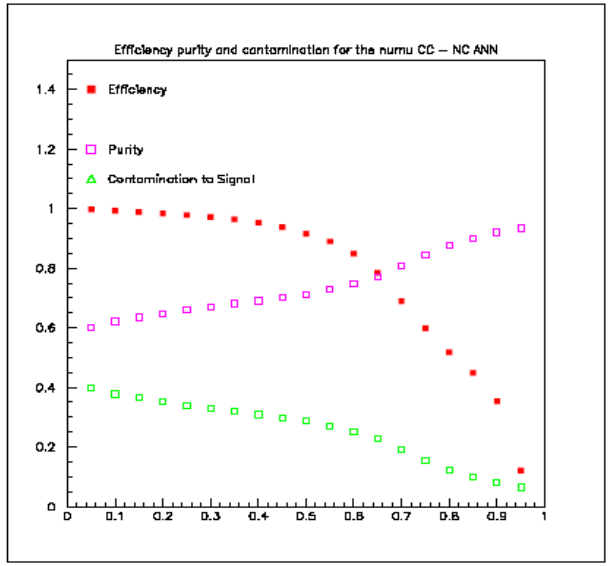
A priori probabilities 1:1.4 Efficiency (red) Purity (magenta) Efficiency x Purity

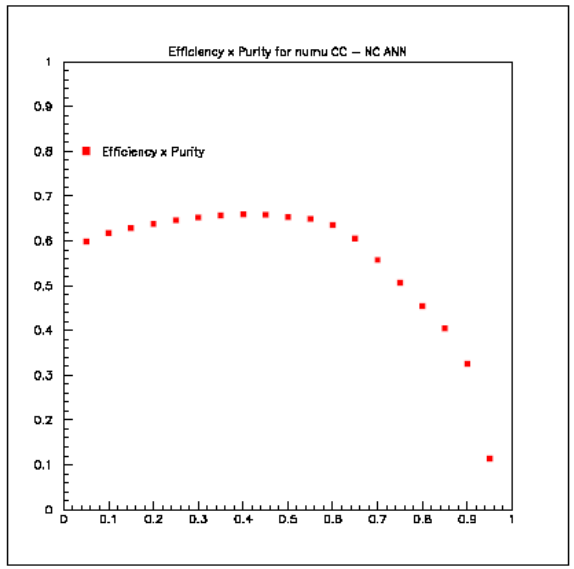
- The results are better & more important different than in Near mostly due to improved S/B ratio (1:1.9 Near vs 1:1.4 Far)
- If we set the cut @ 0.45 (i.e) we have an efficiency of 82% and a purity of 65%.
- As numu;s "oscillate" away the results are going to get better due mainly to the increasing S/B ratio.

ANN Results FAR oscillations (length<40)

$$\sin^2(2\cdot\theta) = 0.95 \;\&\; \Delta m^2 = 0.0025$$







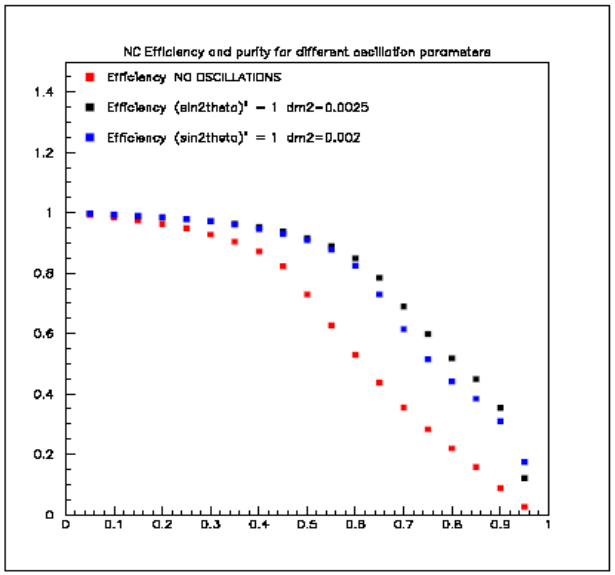
A priori probabilities 1:1.4 Efficiency (red) Purity (magenta)

Efficiency x Purity

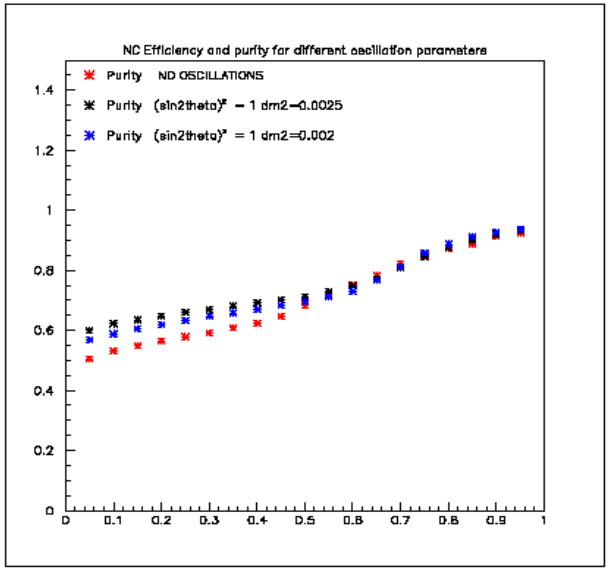
- The results are not that much different than the previous oscillation scenario.
- If we set the cut @ 0.45 (i.e) we have an efficiency of 93% and a purity of 68%.
- The results between Near and Far and furthermore Near and oscillated Far are quite different.

ANN Results FAR :

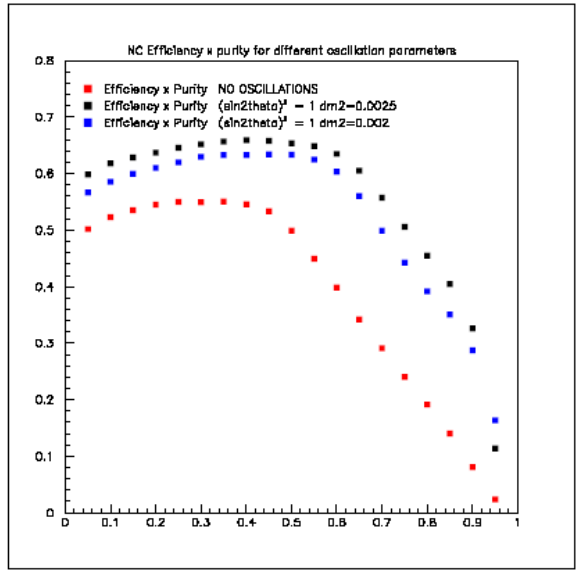
No oscillations - $\sin^2(2\cdot\theta) = 0.95 \,\&\, \Delta m^2 = 0.0025$ $\sin^2(2\cdot\theta) = 0.95 \,\&\, \Delta m^2 = 0.002$



Efficiency



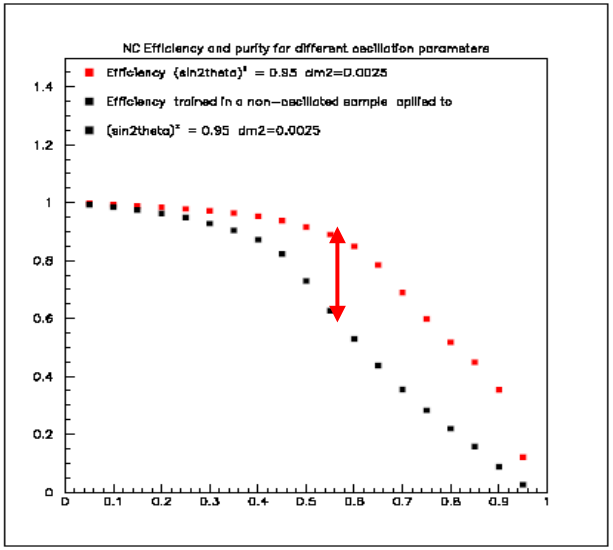
Purity

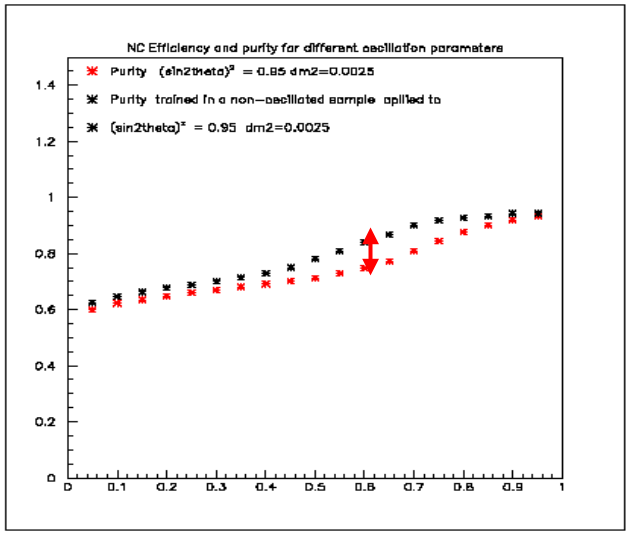


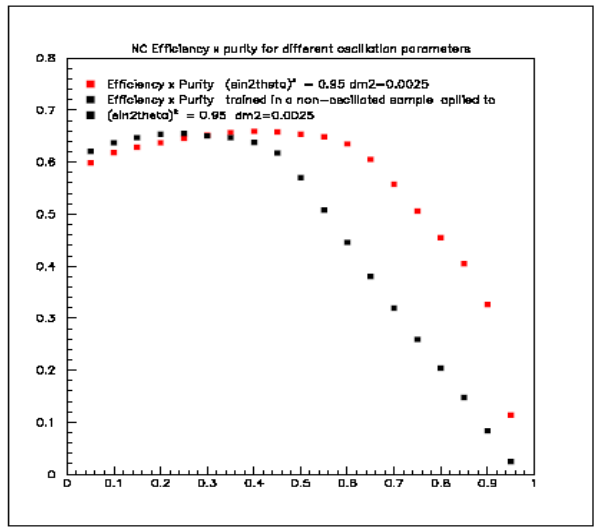
Efficiency x Purity

- Since results with ANN (and any method) truly differ it is important to
  - Train ANN that will be used for the Far detector with Far events
  - Construct maybe a more stable ANN
  - Get the best possible approximation of the oscillation parameters to tune ANN and by doing variations to them estimate the systematic error of the method.

ANN Results FAR con;t



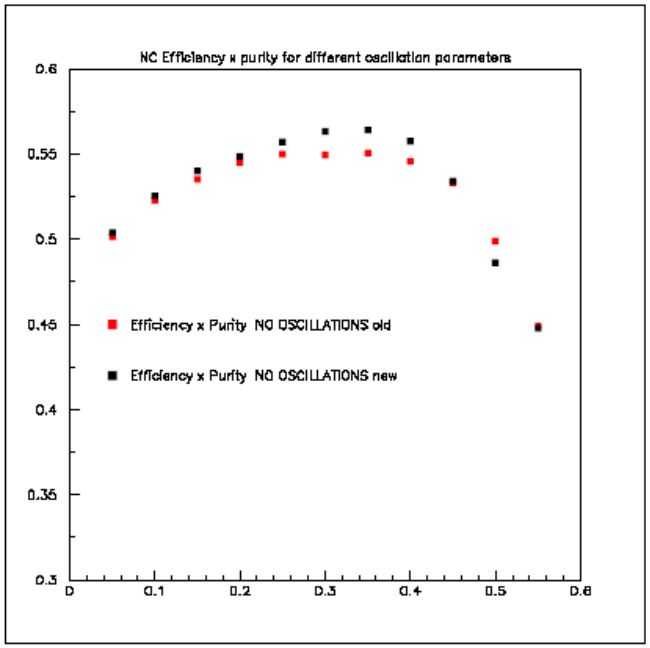


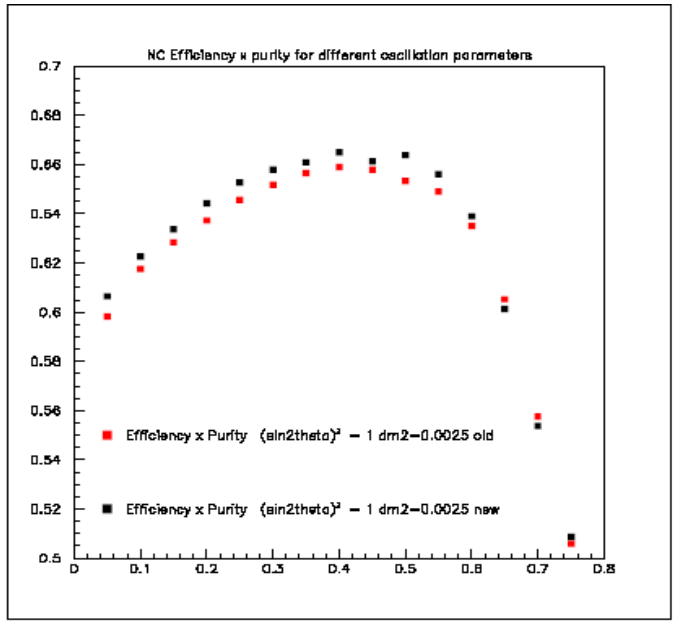


Efficiency **Purity** **Efficiency x Purity**

- The red plots correspond to result obtained if we apply the ANN trained in "unoscillated" events in oscillated ones.
- Efficiency & purity change quite significantly.

ANN Results FAR : Improvement





Efficiency x Purity

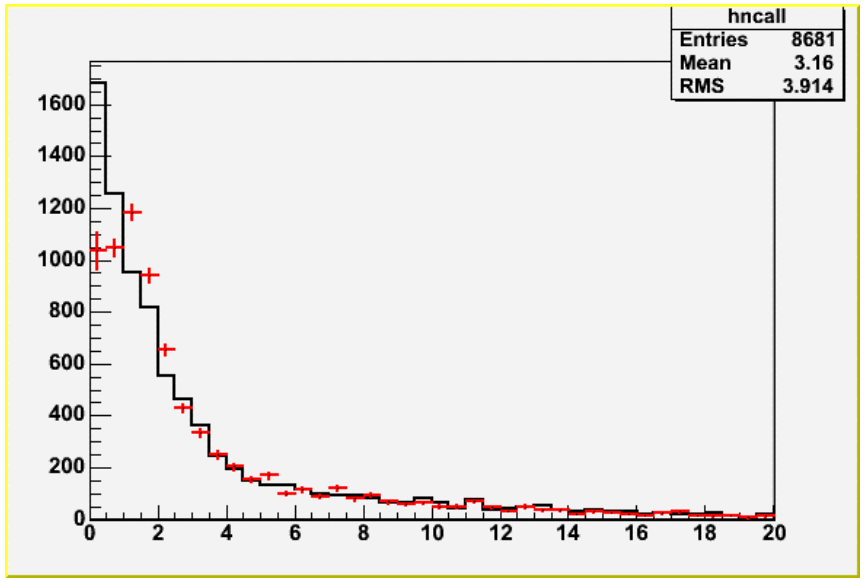
Efficiency x Purity

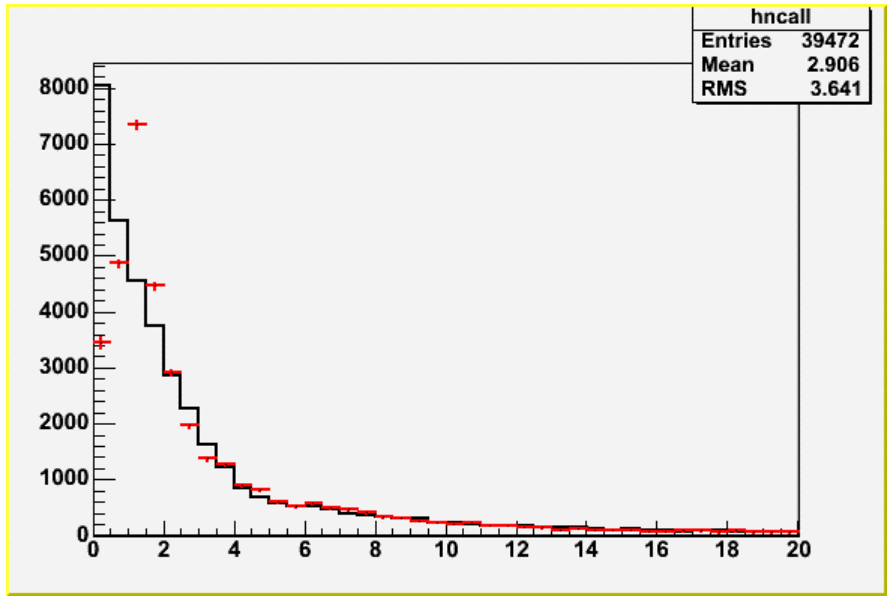
- Using some additional variables correlating tracks with shower there is a slight improvement in both the "unoscillated" events and the "oscillated" ones.
- I expect to see the same behavior in ND as well (results in next meeting)

ANN Results FAR (new) & Near (old in the improving process as done for FAR) : TABLES just for reference & comparison with other methods

cut 0.050 eff 0.994823 pur 0.506438	cut 0.050 eff 0.988004 pur 0.463632
cut 0.100 eff 0.988611 pur 0.531531	cut 0.100 eff 0.972819 pur 0.504156
cut 0.150 eff 0.979737 pur 0.551311	cut 0.150 eff 0.957094 pur 0.530414
cut 0.200 eff 0.966721 pur 0.567460	cut 0.200 eff 0.939213 pur 0.551627
cut 0.250 eff 0.955332 pur 0.582942	cut 0.250 eff 0.919445 pur 0.570580
cut 0.300 eff 0.941725 pur 0.598065	cut 0.300 eff 0.894195 pur 0.586400
cut 0.350 eff 0.922201 pur 0.611574	cut 0.350 eff 0.862971 pur 0.605529
cut 0.400 eff 0.894394 pur 0.623273	cut 0.400 eff 0.826175 pur 0.622512
cut 0.450 eff 0.830498 pur 0.643037	cut 0.450 eff 0.772846 pur 0.644270
cut 0.500 eff 0.718533 pur 0.676413	cut 0.500 eff 0.688382 pur 0.676349
cut 0.550 eff 0.626239 pur 0.715082	cut 0.550 eff 0.589900 pur 0.709308
cut 0.600 eff 0.538678 pur 0.749383	cut 0.600 eff 0.491554 pur 0.746011
cut 0.650 eff 0.443130 pur 0.781022	cut 0.650 eff 0.410998 pur 0.793409
cut 0.700 eff 0.359118 pur 0.815039	cut 0.700 eff 0.354210 pur 0.825118
cut 0.750 eff 0.300399 pur 0.844842	cut 0.750 eff 0.282775 pur 0.852730
cut 0.800 eff 0.230291 pur 0.871780	cut 0.800 eff 0.198535 pur 0.873320
cut 0.850 eff 0.147611 pur 0.891071	cut 0.850 eff 0.111016 pur 0.887253
cut 0.900 eff 0.073066 pur 0.911439	cut 0.900 eff 0.040300 pur 0.888119
cut 0.950 eff 0.020411 pur 0.926175	cut 0.950 eff 0.015770 pur 0.909320

Reconstructed NC spectrum & # of Events in Near & Far





- **Black is true NC spectrum** and red is the estimated after correction for reconstruction efficiency, ANN selection efficiency and ANN purity, and ANN systematic errors.
- The estimated Shower energy is higher than the true shower energy (for reasons discussed previously).
- The obtained NC/ALL ratio is: (for $\sin^2(2\cdot\theta) = 0.95 \,\&\, \Delta m^2 = 0.0025$)
  - Near estimated : 25.1% ± 0.1% Near True : 25.2 %± 0.1%
  - Far estimated : 33.8% ± 0.6% Far True : 33.8% ± 0.6%

Summary / On going work

- Separating NC events using Neural Networks is a quite successful and powerful method & with the addition of new variables I achieved slight improvement. Similar work for NEAR is underway.
- The method of reconstructing the shower energy (and neutrino energies for CC) events is quite promising, has the advantage that makes the energy resolution much better and more similar between Near and Far and I already started working towards additional improvements and more detailed work on the problem.
- The ANN for NC/CC separation and the ANN shower energy are parameterized functions in C (or C++) and I plane to create an Analysis Module that will read the needed event information from the Ntuples and produce for each Event its PID and Estimated shower energy. More on that next time…