status of heavy flavour tagging

S.Xella Hansen

Very rich program, to carry out before september conference Vertex 2001 and ECFA-DESY studies workshop (some of it already for Snowmass)

- validating the tagging procedure on ee$\rightarrow$qq events at 91 GeV. Some things puzzling left over from Tesla TDR studies (e.g. imp. parameter distribution for uds events), so need a carefully study of input variables (and algorithms) to NN, and consistency checks on NN.

- find areas of possible improvement (e.g. ghost algorithm, join probability distributions retuning, ...)

- look at 400 GeV and 800 GeV energy regime, in physics case of ee$\rightarrow$ qq, and study performances (purity vs efficiency) there too.

- start looking at other performances (e.g. vertex charge)

- finally, apply all this to a more interesting physics case

People working actively on this for LCFI at the moment: myself, Matthew Wing (Bristol).

More recently, contact with Arnaud Gay (Mark Winter postdoc) about impact parmeter joint probability tag, with Andre Sopzcak (Lancaster) concerning new ideas for tagging, and Ivanka Bozovic-Jelisavcic (collab. with Klaus Moenig at DESY-Zeuthen) concerning gamma gamma collider physics.

tagging validation

tagging is done using NN. several inputs:

- ZVTOP: mass, pt, n of vertices, etc..
  present variables and algorithm carefully studied for Tesla TDR, so we are quite confident here.
  IMPROVEMENTS: ghost track algorithm , which will improve vertex finding efficiency and jet flavour/anti-flavour separation

  Timescale for this: september more realistic

- IMP. PARAMETER JOINT PROB. : product of probabilities (r,phi), (r,z), (3d) for tracks in a jet to come from the primary vertex.
  IMPROVEMENTS: A bit of work must be done here, to re-estimate the probability distribution to be used in the algorithm for the different energies, plus try to release cuts in the tracks used by the algorithm (now only tracks with imp. param. significance lower than 50, not very sensible)
  This input to the tagging gives some higher performance: e.g. b tagging, on events with 2 vertices found: FoM of NN goes from 0.5937 +/- 0.0029 to 0.6109 /- 0.0028 by introducing this probability information . It might also improve further the charm tagging.

  Timescale for this: by end of next week done for all energies (91,400,800 GeV)

- 1-PRONG: highest $\frac{d0}{\sigma(d0)}$, $\frac{z0}{\sigma(z0)}$, ... in a jet used to recover those events where no secondary vertex could be found (class 0-1 vtx: more important for c tagging).
  need to understand the distribution of these variables for uds and charm jets + find other usefull inputs for NN.

  STATUS:

  – We think we understand the reason for tails in uds distribution, caused by decays of b/c generated in gluon splitting and by weak decays of starnge barions.

  – momentum and angle wrt jet of highest i.p. track introduced in NN $\rightarrow$ slight improvement:
    b net 0.2411 $\rightarrow$ 0.2383 +-0.0119
    c net 0.1121 $\rightarrow$ 0.1152+-0.0079
    bc net 0.1299 $\rightarrow$ 0.1617+-0.0075

  – sanity checks being performed: not using the information on imp. param. significances for charm tagging pushes the performance curve eff. vs pur. at lower efficiencies.
    not considering events with only IP reconstrcuted, pushes charm curve towards b/c curve , which is reasonable (no uds background)
    More...

  So this tagging method seems quite reasonable now.

  IMPROVEMENTS: finish some further sanity checks.

  Timescale for this: end of next week final checks done.

Running tagging at higher energies

Started. Need mainly re-estimation of prob. distribution for joint prob. tag.

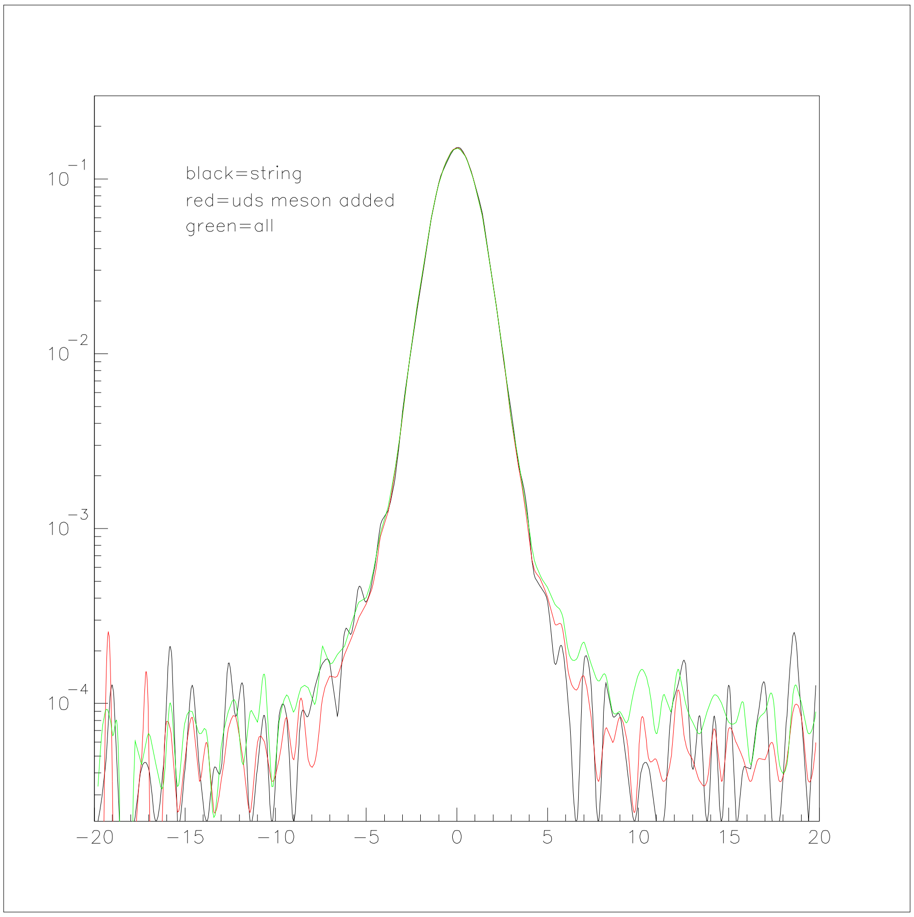
Plan is to have by end of next week results for 400 and 800 GeV purity vs efficiency. Thanks to Matthew for the good work done here, thanks to him we can look at this energy regime with our simulation now!

Why is this important?

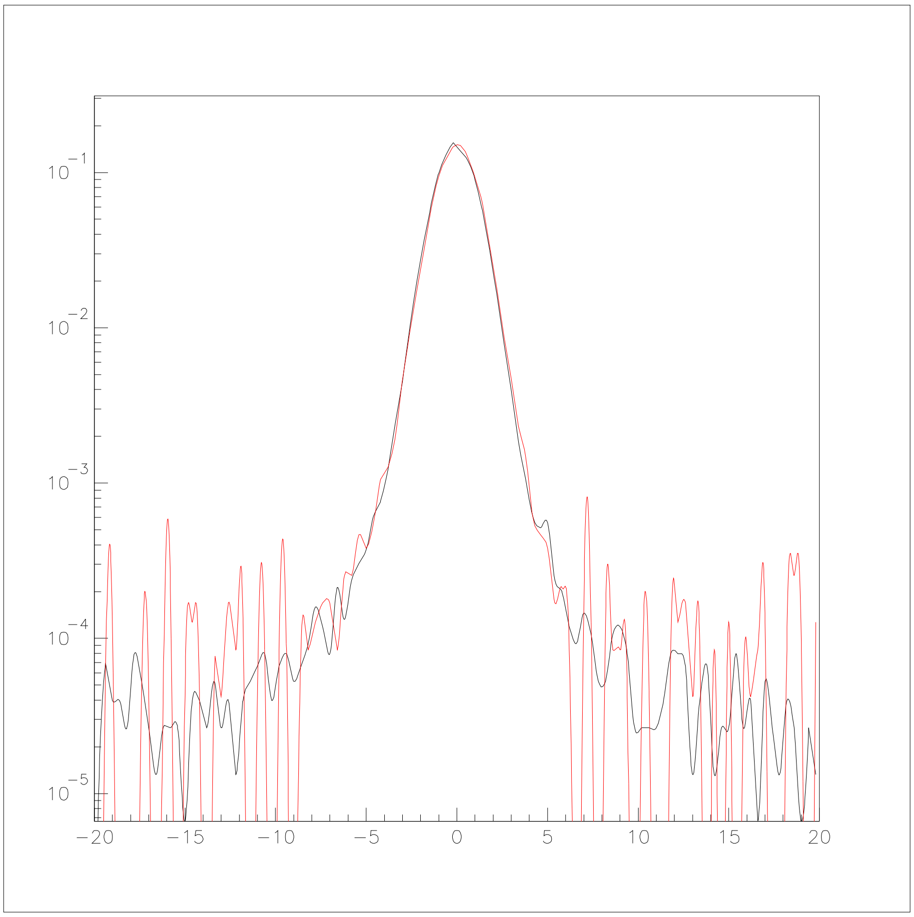
We have not tested what happens to tagging when energies goes up in a realistic physics case (like qq bar events). Last thing done was with monojets.

If we are sure we understand behaviour of tagging on this benchmark process, then we can safely use it on more interesting physics processes for LC.

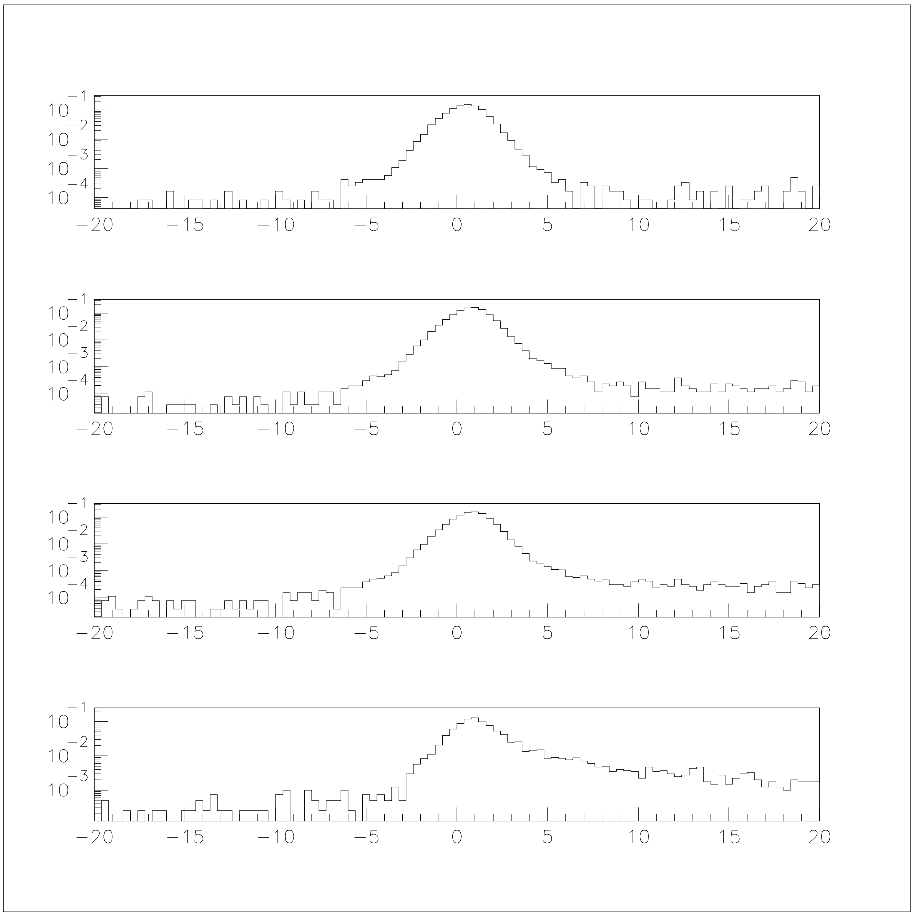
tracks in uds jets, splitted in components according to their mother: string(black), uds mesons like ρ,etc..(red), all (so including strange baryon or B/D from gluon splitting) (green). Distributions are normalized to their number of entries. x=imp. param. significance r,phi



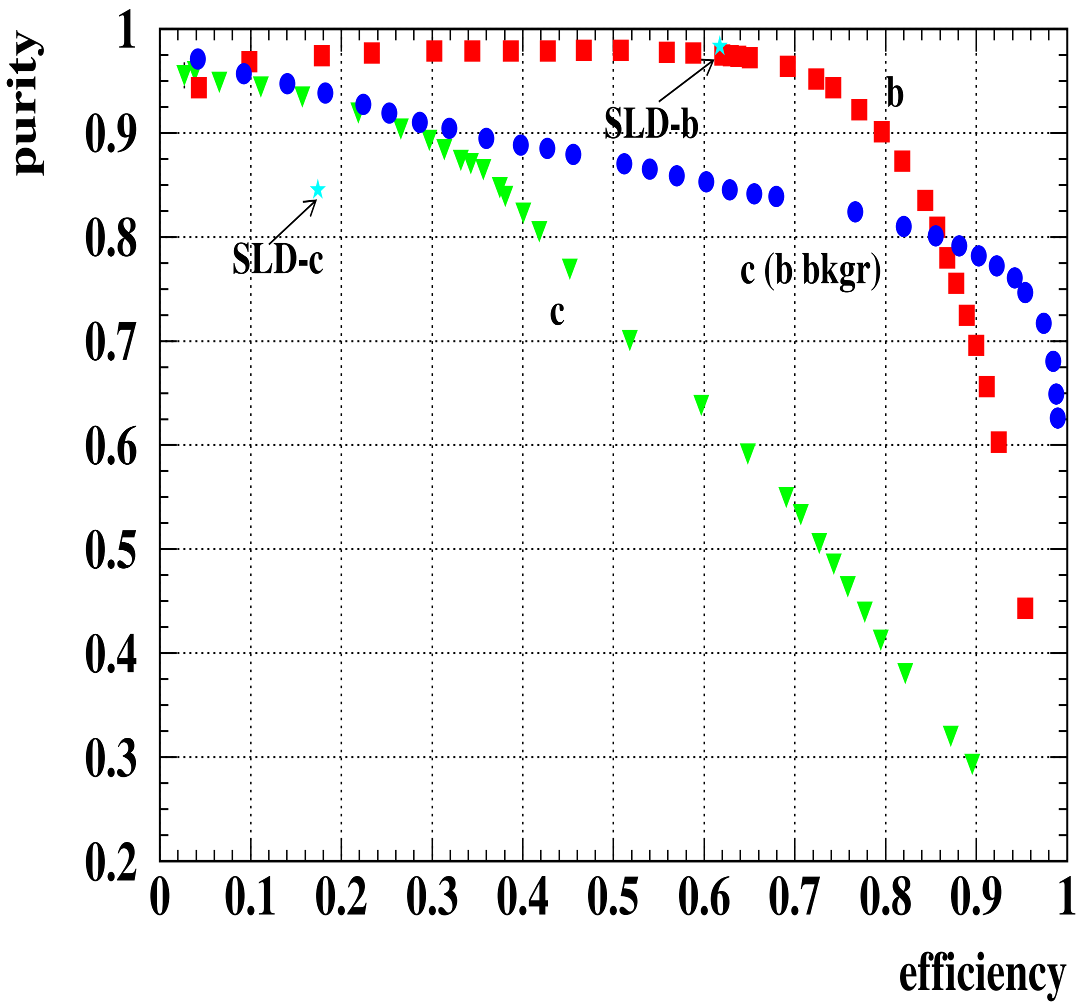
tracks in uds jets coming from string (red), compared to single pions sent through detector (black). Distributions are normalized to their number of entries. x=imp. param. significance r,phi



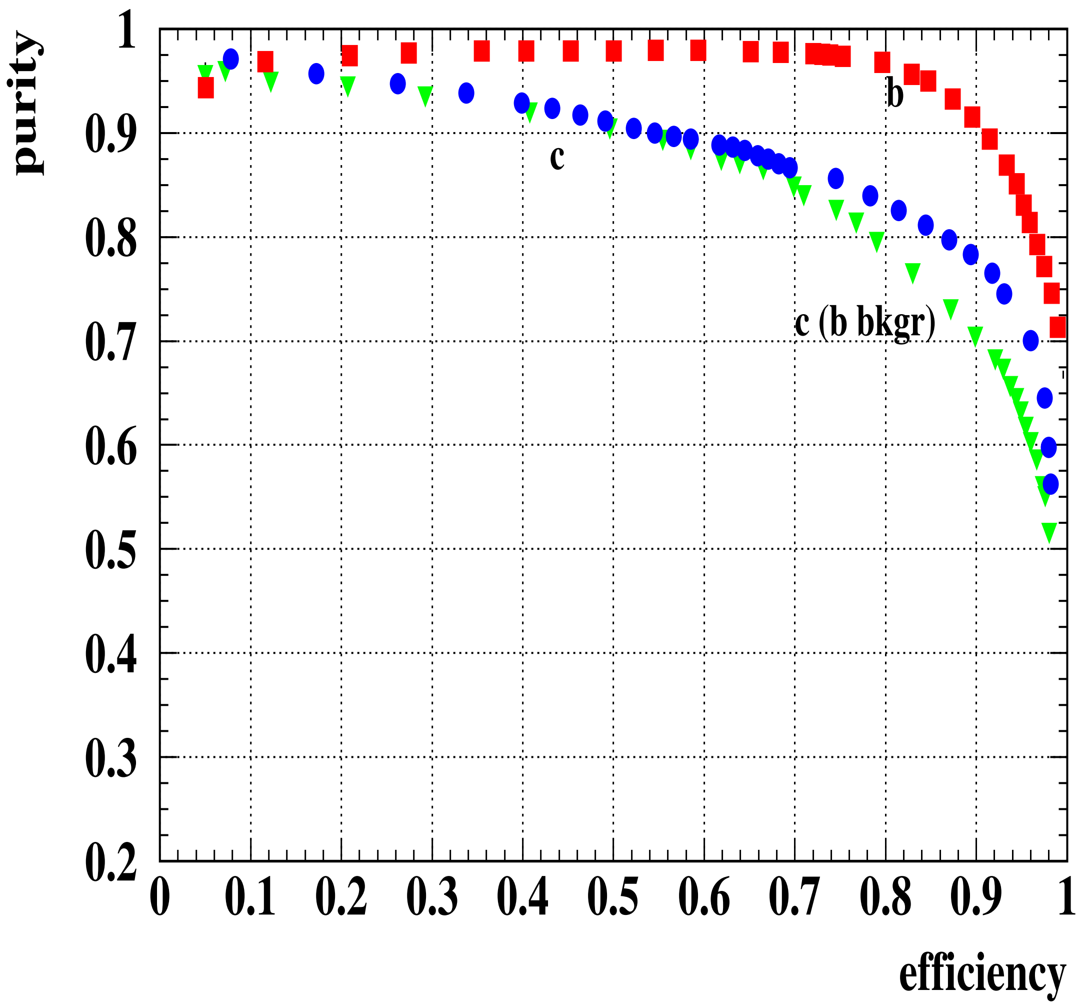
from top to bottom: highest imp. parameter track in uds jets coming from string, from string and uds meson, from all (incl. strange baryon or B/D), and finally in ccbar jets.Distributions are normalized to their number of entries. x=imp. param. significance r,phi



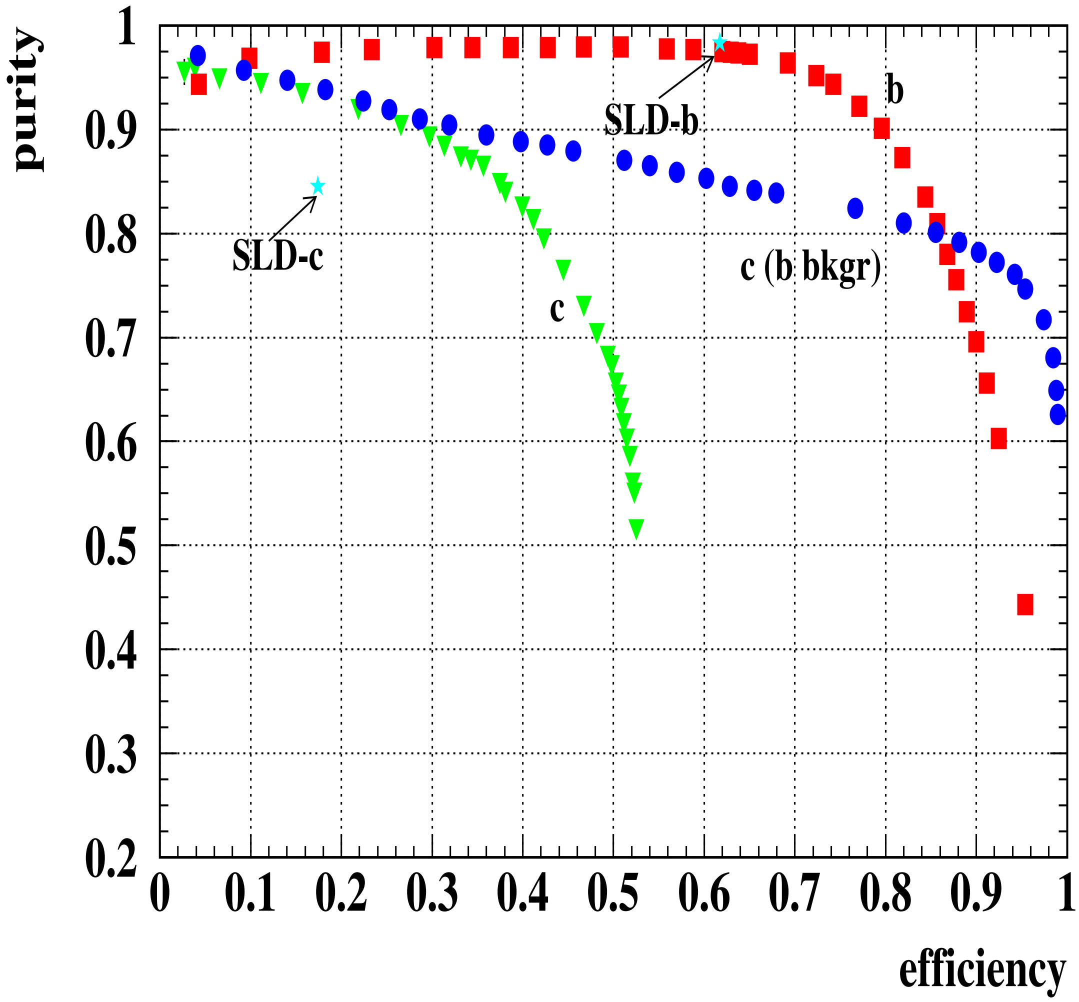
Current status of tagging performance on 91 GeV ee→qq events.
B field is 4 Tesla



NN performance on 91 GeV ee→qq events, not using at all events with only 1 vtx (IP) reconstructed.B field is 4 Tesla

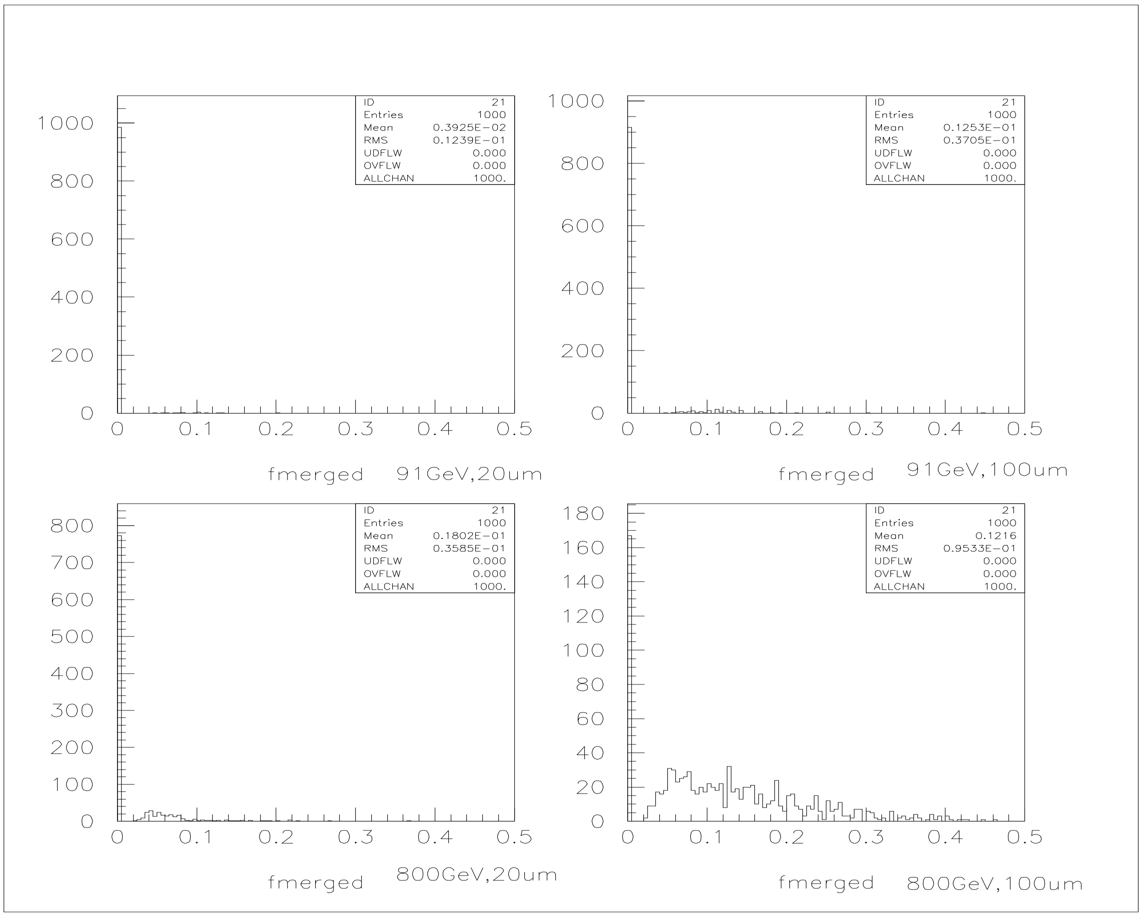


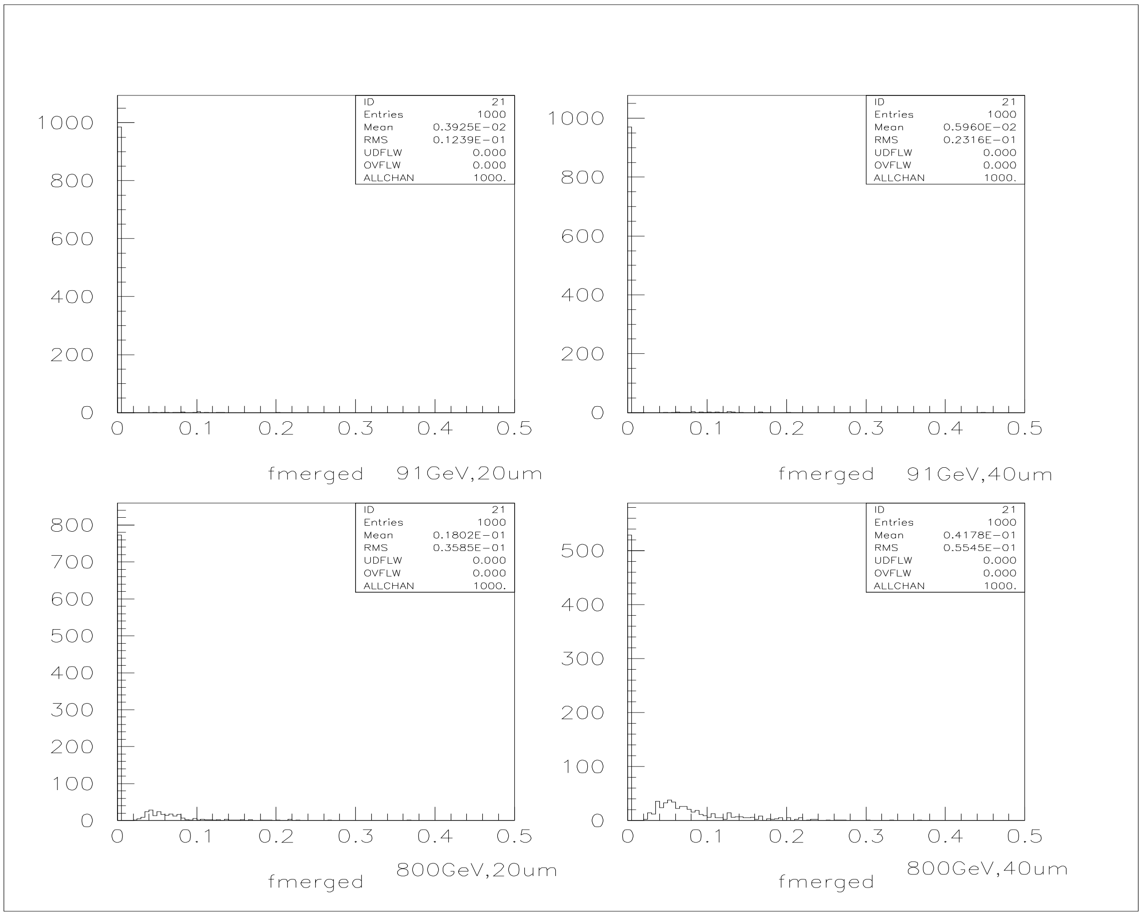
NN performance on 91 GeV ee→qq events,for charm tagging not using informations from max imp param significant track in jet. B field is 4 Tesla



Appendix A: study of merging of tracks at different energies and pixel sizes

Here we study the merging of tracks at different energies and pixel sizes. We make the assumption that if a track hits a pixel, it will also leave signal in the 4 adjacent pixels. In the following 2 plots, with fmerged we indicate the fraction of tracks in layer 1 which have at least one pixel in common with other tracks. Number of events is 1000 for each plot, and 1 entry for each event. Events are of type ee $\rightarrow$ qq .





ID 21
Entries 1000
Mean 0.3925E−02
RMS 0.1239E−01
UDFLW 0.000
OVFLW 0.000
ALLCHAN 1000.
1000
800
600
400
200
0
0 0.1 0.2 0.3 0.4 0.5
fmerged 91GeV,20um
ID 21
Entries 1000
Mean 0.5960E−02
RMS 0.2316E−01
UDFLW 0.000
OVFLW 0.000
ALLCHAN 1000.
1000
800
600
400
200
0
0 0.1 0.2 0.3 0.4 0.5
fmerged 91GeV,40um
ID 21
Entries 1000
Mean 0.1802E−01
RMS 0.3585E−01
UDFLW 0.000
OVFLW 0.000
ALLCHAN 1000.
800
700
600
500
400
300
200
100
0
0 0.1 0.2 0.3 0.4 0.5
fmerged 800GeV,20um
ID 21
Entries 1000
Mean 0.4178E−01
RMS 0.5545E−01
UDFLW 0.000
OVFLW 0.000
ALLCHAN 1000.
500
400
300
200
100
0
0 0.1 0.2 0.3 0.4 0.5
fmerged 800GeV,40um

Appendix B: study of minimal distance of track pairs on ly1 for different energies

Here we study the minimal distance of track pairs on layer 1 for different energies. In the following plot, for a given value of minimal distance (in micrometers), the fraction of events having a minimal distance for pairs on ly1 smaller than that is plotted. 3 different energies are considered. Events are of type ee $\rightarrow$ qq .