TILC08

Joint ACFA Physics and Detector Workshop and GDE meeting

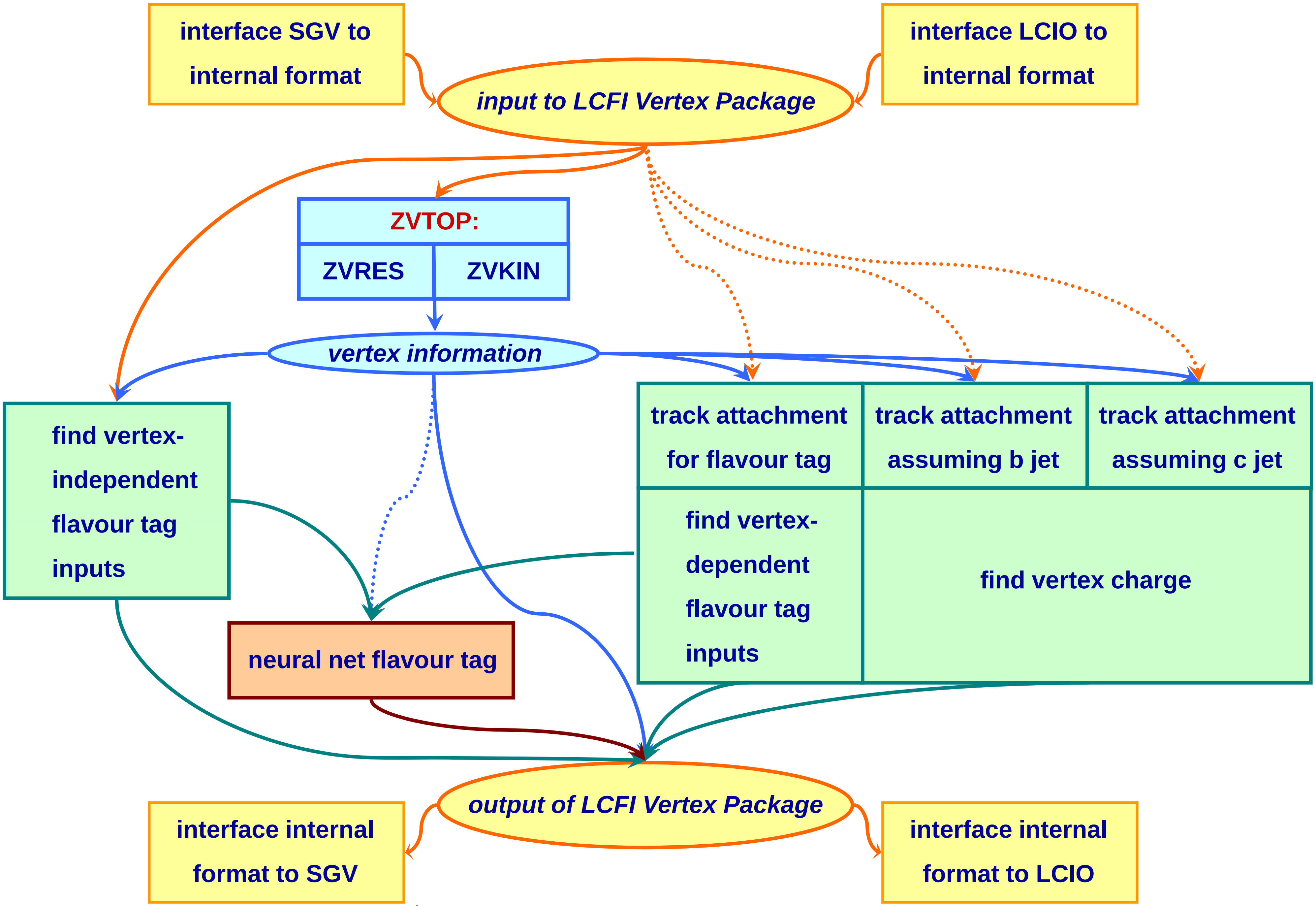
on the International Linear Collider

3 – 7 March 2008

LCFI Vertex Package: Parameter Optimisation

- Introduction
- Track selection for full MC and reconstruction
- Status of parameter tuning
- Quark charge quality indicator

Sonja Hillert (Oxford) on behalf of the LCFI collaboration



interface SGV to internal format
interface LCIO to internal format
input to LCFI Vertex Package
ZVTOP:
ZVRES
ZVKIN
vertex information
find vertex-independent flavour tag inputs
track attachment for flavour tag
track attachment assuming b jet
track attachment assuming c jet
find vertex-dependent flavour tag inputs
find vertex charge
neural net flavour tag
output of LCFI Vertex Package
interface internal format to SGV
interface internal format to LCIO

Preparation for mass reconstruction - overview

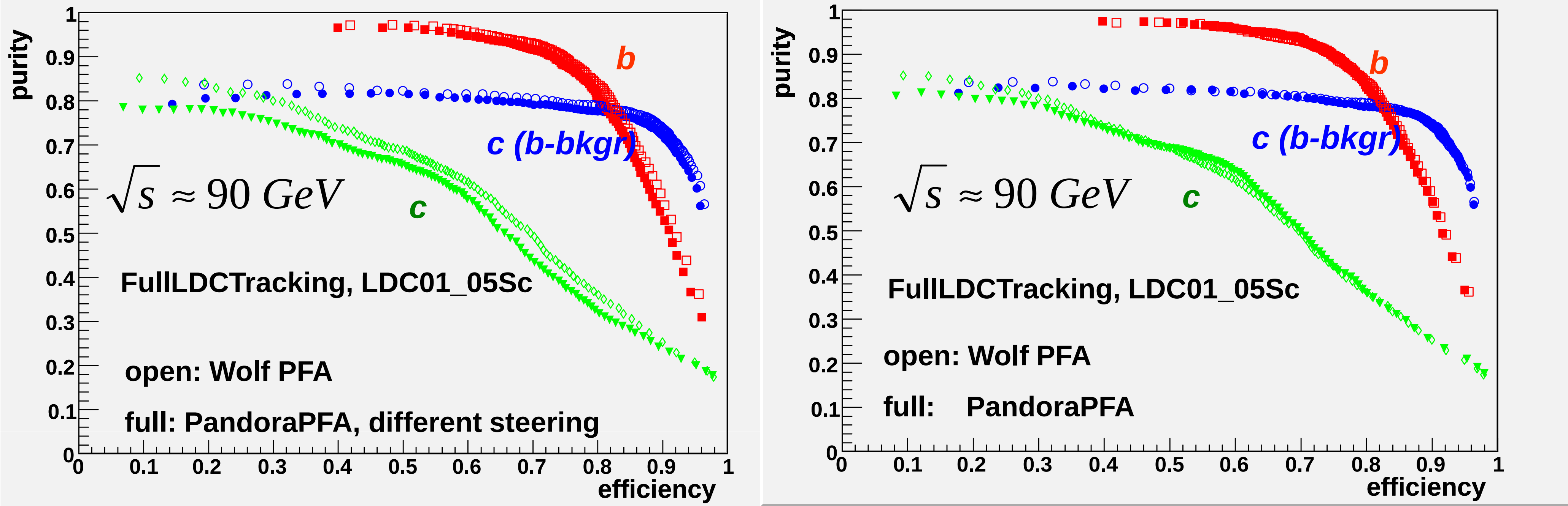
- **ensure compatibility of LDCtracking, PandoraPFA, LCFIVertex** **DONE** ***K Harder et al. (RAL)***
- **check LCFI code runs without problems with the new detector models (VXD03, VXD04, new SiLC detectors when available)** **ONGOING** ***Clare Lynch (Bristol)***
- **implementation of correction procedures based on reconstructed information:**
  - **photon conversions**
  - **K_S and Λ decay tracks**
  - **hadronic interactions**

  FIRST RESULTS ***Kristian Harder (RAL)***
- **parameter tuning with full MC and reconstruction:**
  - **optimise track selection for ZVRES and IP-fit based**
  - **tune ZVRES parameters**

  ONGOING ***Roberval Walsh (Edinburgh)***
  - **optimise track selection for flavour tag**
  - **tune flavour tag parameters, in particular: parameters for joint probability calculation**
- **training of new neural nets using full MC, reconstruction and optimised code parameters** **IN PREPARATION** ***Roberval Walsh (Edinburgh)***

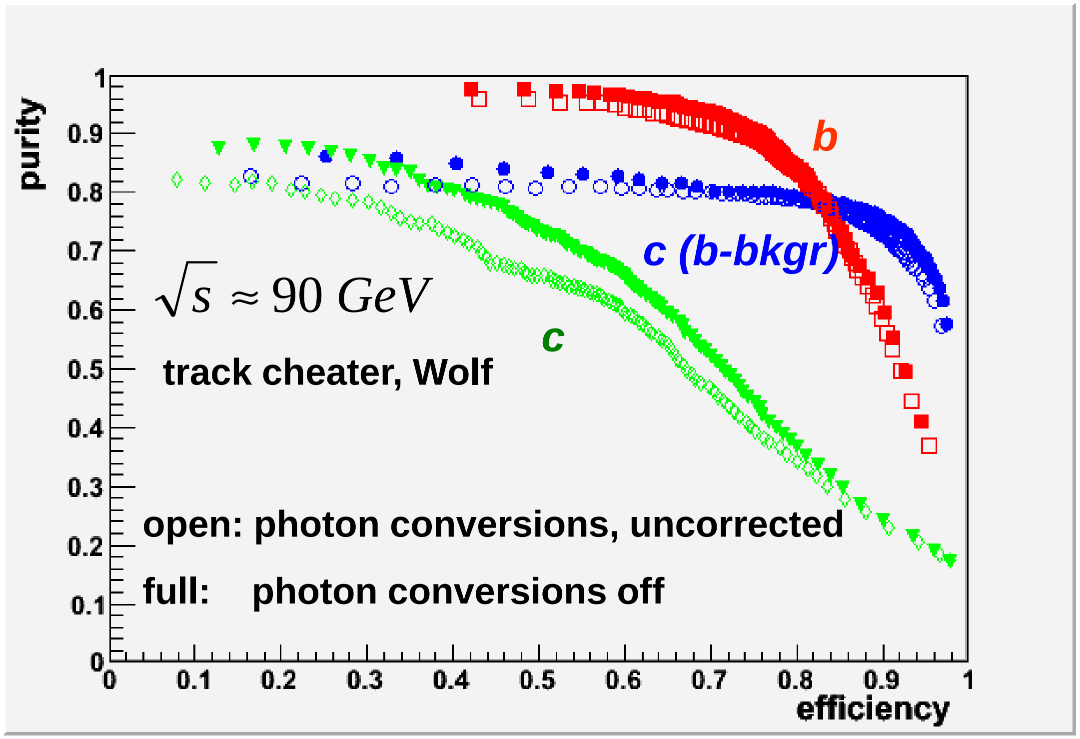
Compatibility with other packages

Kristian Harder, Talini Pinto Jayawardena (RAL)



- The initial result (left, ILD optimisation meeting 30th January) seemed to indicate worse flavour tag performance with PandoraPFA input than with Wolf input
- This was traced back to differences in the FullLDCTracking steering
- Using same tracking settings, performance is in excellent agreement
- Mark Thomson has checked that PFA performance is not affected (ILD mtg 13th Feb)

Correction procedures (conversions, K_S, Λ)



- during validation of LCFI code, photon conversions switched off at GEANT-level
- Central switch that also affects CAL showers → NOT VIABLE FOR MASS PRODUCTION
- so far, K_S and Λ decay tracks removed using MC information

- 2 new Marlin processors in preparation *(Kristian Harder, RAL)*:
  - standalone conversion / V0 tagger
  - performance analysis for standalone tagger

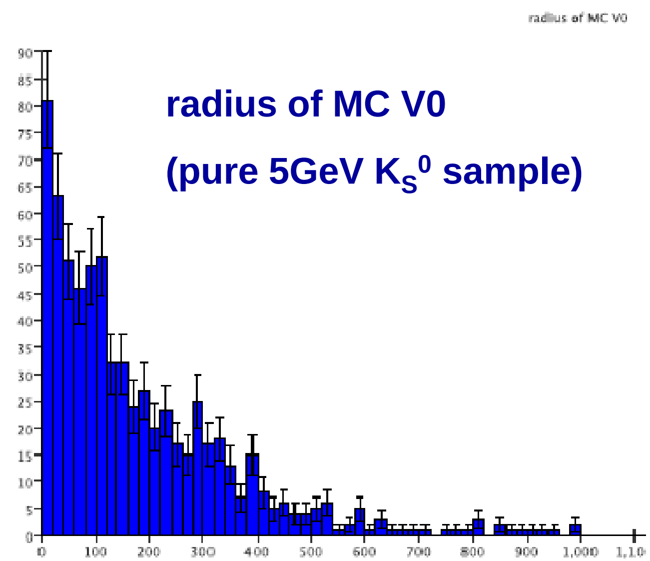
Conversion tagger performance analysis : first results

New processor to monitor performance

- **scans all MCParticle collections for actual conversions/V0**
- **scans all Track collections for matching reconstructed tracks**
- **scans all ReconstructedParticle collections for matching objects:**
  - **either individual track objects**
  - **or composite objects involving conversion/V0 tracks**
- **output mostly in the form of tables + a few histograms: looks reasonable**

```
PERFORMANCE SUMMARY:
actual conv/V0 in MCParticles: 774 with 1558 tracks
TrueTracks collection contains 1727 tracks,
   including 1429 ones actually coming from V0/conversion
   ==> efficiency 0.917202
   ==> purity     0.827446
TrueTracks collection contains 0 correct tracks in composite objects
   and 0 tracks in composite objects with >1 track
   ==> composite efficiency 0
TrueTracksConv collection contains 2104 tracks,
   including 1413 ones actually coming from V0/conversion
   ==> efficiency 0.906932
   ==> purity     0.671578
TrueTracksConv collection contains 1413 correct tracks in composite objects
   and 2104 tracks in composite objects with >1 track
   ==> composite efficiency 0.906932
   ==> composite purity     0.671578
```

(Kristian Harder, RAL)



Parameter tuning: approach

➢ **A lot of progress has been made preparing this complex study *(Roberval Walsh)*:**

- **ensured MarlinReco runs on Edinburgh cluster (1400 nodes, 64-bit machines) and gives reasonable results (runs 3 times faster than on local machine)**

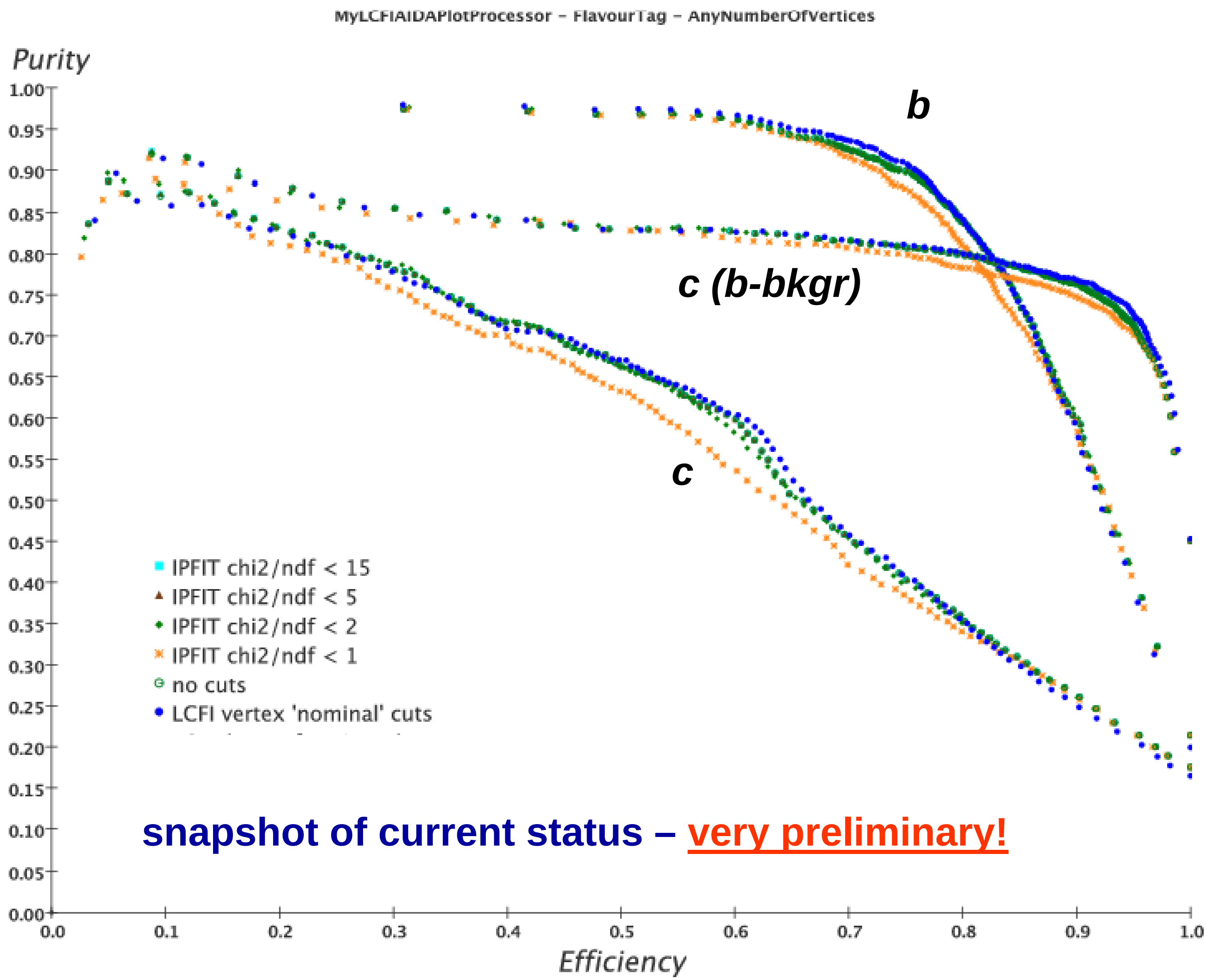
- **main criterion for optimisation of parameters: flavour tag performance in terms of purity vs efficiency; parameters varied one at a time & performance evaluated**

- **scripts written and checked to automatically vary parameters under study in steering files**

- **First results obtained for track selection for IP fit**

- **Production of a training sample for flavour tag NN's underway on the GRID: Many thanks to the DESY-group!**

Example: effect of varying track χ^2/ndf cut for IPfit



Flavour tag: parameters for joint probability

- improvement to calculation of one of the inputs for the flavour tag-NN:

- probability that a track from primary vertex has impact parameter significance > b/σ_b is

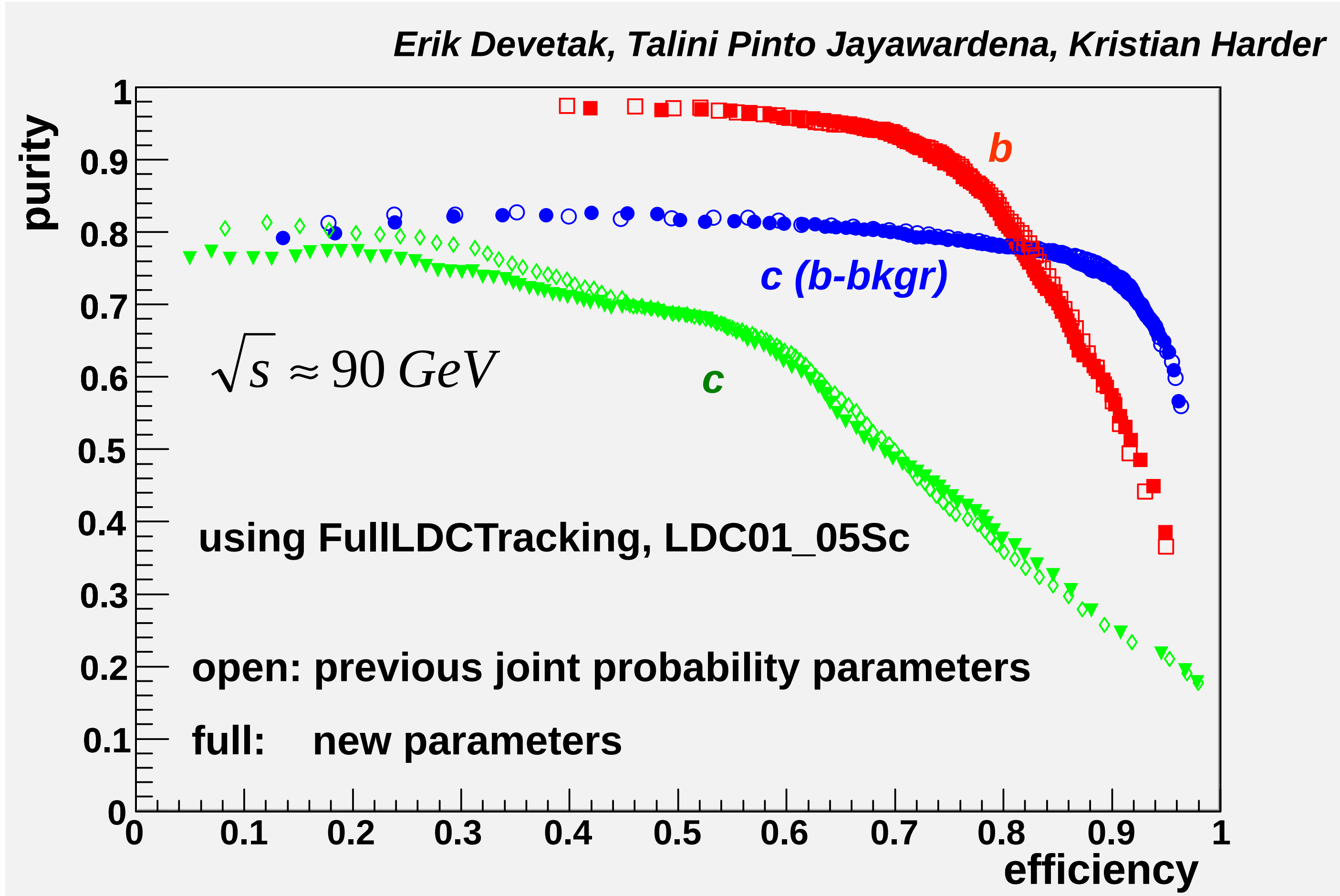
$$P_i = \frac{\int_{b/\sigma_b}^{\infty} f(x)\,dx}{\int_0^{\infty} f(x)\,dx}$$

- the symmetric function $f(x)$ needs to be obtained from a fit of the negative side of the impact parameter distribution (done separately for R-ϕ and z);
- current version of LCFI Vertex uses hard coded parameterisation obtained from fast MC
- Erik Devetak implemented and tested a module to obtain parameters properly from a fit

Fit macro for joint probability calculation: results

parameters for Rϕ joint probability	
previous	new
1.013	1.015
0.025	0.280
0.102	0.561
0.041	0.006
0.016	0.048

parameters for Z joint probability	
previous	new
1.016	1.043
0.027	0.271
0.095	0.468
0.041	0.006
0.015	0.048



Obtained new values with full MC & reconstruction (left):

values change as expected, flavour tag performance stable.

Quark charge quality indicator: motivation

➢ Idea: find a quality indicator for the quark charge – similar to the flavour tag:
The higher the value of the quality indicator, the more likely the quark charge has been correctly reconstructed

➢ In analyses, this could be used as a kind of "quark charge tag":

- to resolve ambiguities in events with 2 or more b-jets where results are ambiguous
  (e.g. events with 2 b-jets both with same reconstructed quark charge,
  or for events with 4 b-jets, where one knows quark charge is wrong for >= 1 jet(s))
- to improve the event selection, yielding a sample with better quark charge purity

➢ I had a first look at the possibilities for this with SGV (need a lot of statistics!):

- simple quality indicator defined
- studied resulting improvement in quark charge selection for 2-jet events (Z-peak)

Approach

- For initial study, obtain quark charge purity as function of a number of sensitive variables (histogram correctly & wrongly found quark charge, calculate bin-wise correct/total)
- Distributions look considerably different for two- and three (or more) vertex jets

→ Need to do this separately for the two categories (as for flavour tag)

- For each jet within jet-selection cuts, calculate a quality estimator as follows: multiply individual purities $\pi(v_i)$ for each of the input variables v_i:

$$Q_{quark}\text{-tag} = \Pi_{\,i = 1..n}\ \pi(v_i)$$

- For sample of two-jet events at the Z-peak, find quark charge leakage rate,

$$\lambda_q = N(Q_q \text{ wrong}) / N(\text{jets in sample})$$

for the usual quark charge reconstruction and an improved procedure (details later)

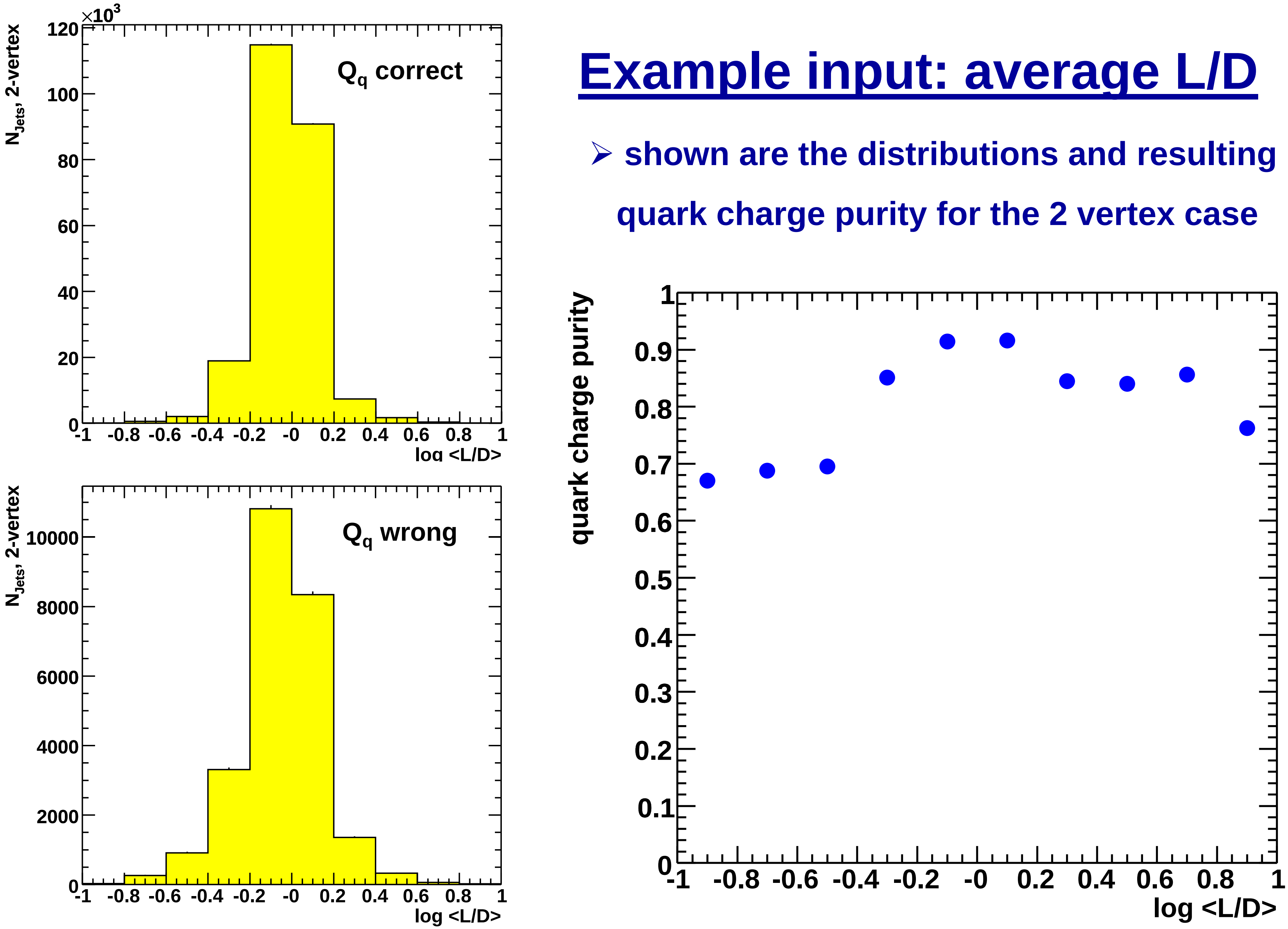
Input variables used

- **As input variables used some of the flavour tag variables, namely:**
  - **decay length and decay length significance of most significant vertex,**
  - **M_{Pt}**
  - **seed vertex momentum**
  - **sum of number of tracks in all non-primary vertices**

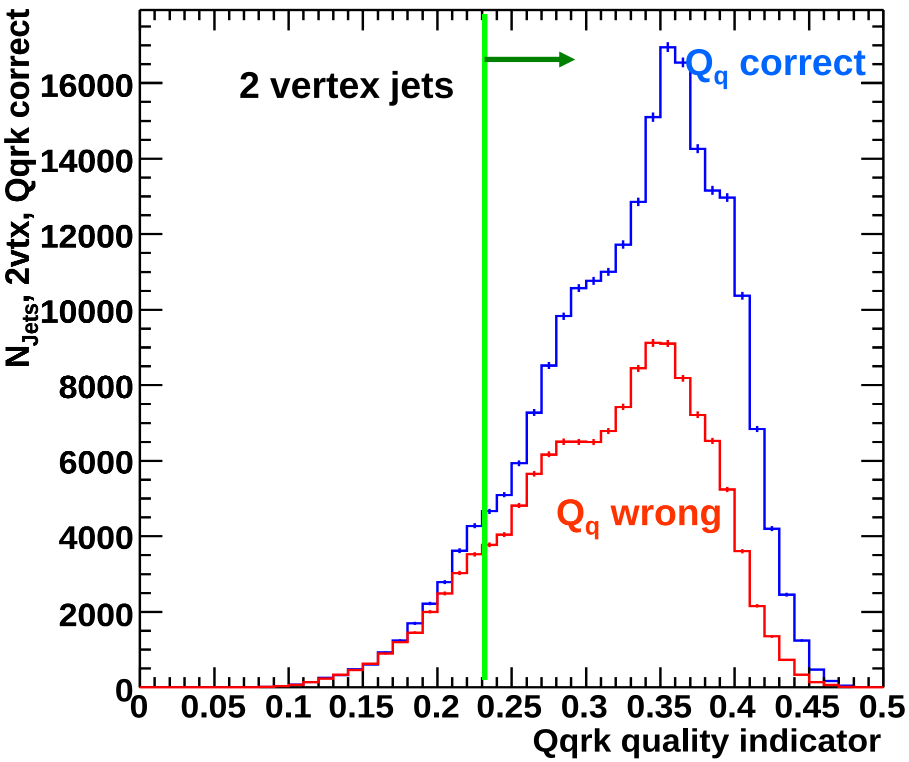
- **Further variables included (should be specifically sensitive to vertex & quark charge quality):**
  - **<L/D>: average of L/D value for all tracks assigned to decay chain**
  - **$(L/D)_{max}$, $(L/D)_{min}$: maximum and minimum values of L/D of assigned tracks**
  - **$(L/D)_{max} - (L/D)_{min}$**
  - **<T>: average T variable (measure of transverse distance of track from seed axis)**
  - **T_{max}, T_{min}: maximum and minimum values of T for assigned tracks**
  - **$T_{max} - T_{min}$**

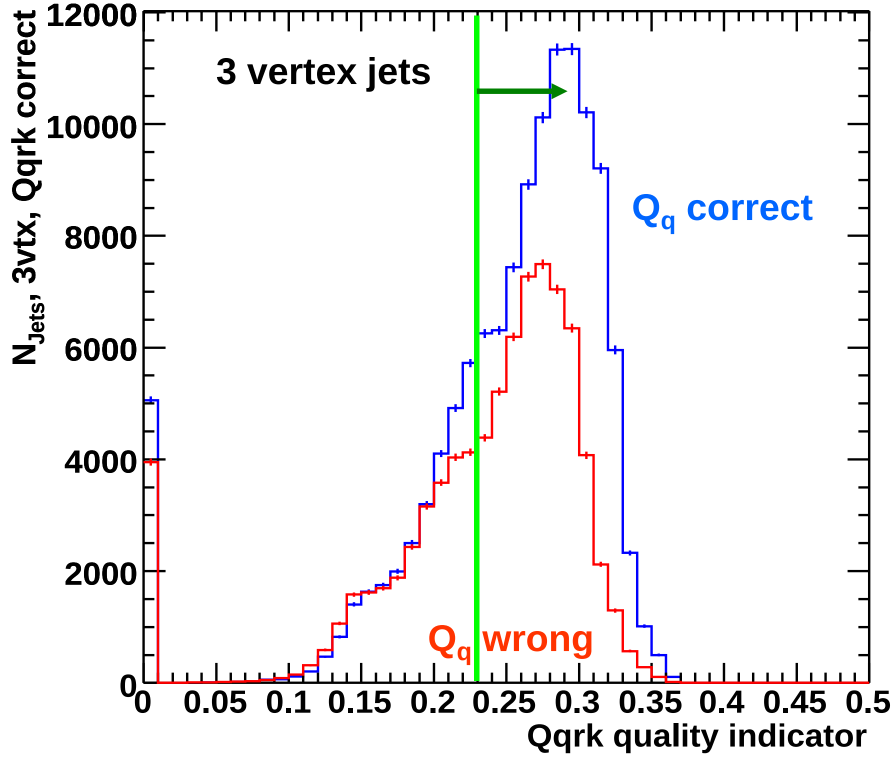
Example input: average L/D

➢ shown are the distributions and resulting quark charge purity for the 2 vertex case



Improved quark charge determination for 2-jet events





➢ **Example how this can be used to improve quark charge in 2-jet events:**

- **for jets with vertex charge 0, can use information from other jet, if good quality (see cut above)**
- **if same reconstructed quark charge is found for both jets, use the one with higher quality estimate, after subtracting an offset taking into account that the 2-vertex and the 3-vertex distribution peak at different values**

Results from improved correction procedure

➢ For clash category: before comparing quality estimators, subtract 0.35 for jets with 2-vertices, 0.28 for jets with 3 or more vertices

	raw Q_q correct	raw Q_q wrong	modified Q_q correct	modified Q_q wrong
opposite sign	~ 123700	2674	~ 123700	2674
this jet 0	0	~180500	~ 112000	7861
other jet 0	~ 210300	~ 23100	~ 210300	~ 23100
clash	~25300	~25300	~34400	~16100

	raw Q_q	modified Q_q
opposite sign	2.1 %	2.1 %
this jet 0	100 %	6.6 %
other jet 0	9.9 %	9.9 %
clash	38.0 %	25.3 %

- ~ 60600 “this jet 0” jets discarded
- increase in efficiency by 30%
- resulting quark leakage for entire (new selection) sample: 8.2%

Next step: train NN with SGV, implement NN-based Marlin processor using this net

Summary

➢ LCFI Vertex package works reliably with both Wolf and PandoraPFA;

➢ Performance with new vertex detector geometries VXD03, VXD04 being checked

➢ Correction procedures for photon conversions, K_S and Λ decay tracks and hadronic interactions being implemented, first results look promising

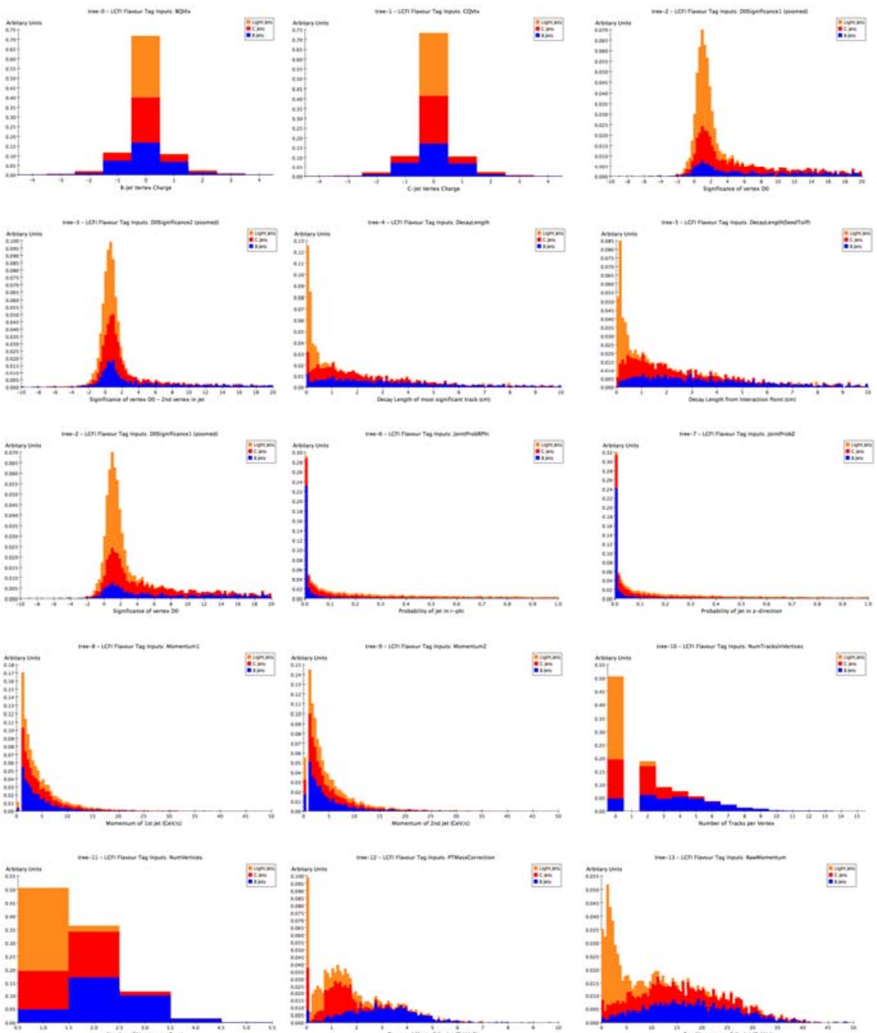
➢ Preparation for code parameter tuning making good progress: semi-automatic scripts written, samples generated for retraining of neural networks with full MC & reconstruction

➢ Quark charge "tag" permits improvement of quark charge sign selection:

- Simple quark charge quality indicator implemented and tested using 2-jet events at the Z-peak
- At a b-tag efficiency of 70%, by combining information from both jets in the event can increase efficiency by 30% and achieve quark charge leakage of 8.2% (i.e. 1.6 % less than with standard procedure); neural net-based approach to be implemented in Marlin shortly

Additional Material

Diagnostic features



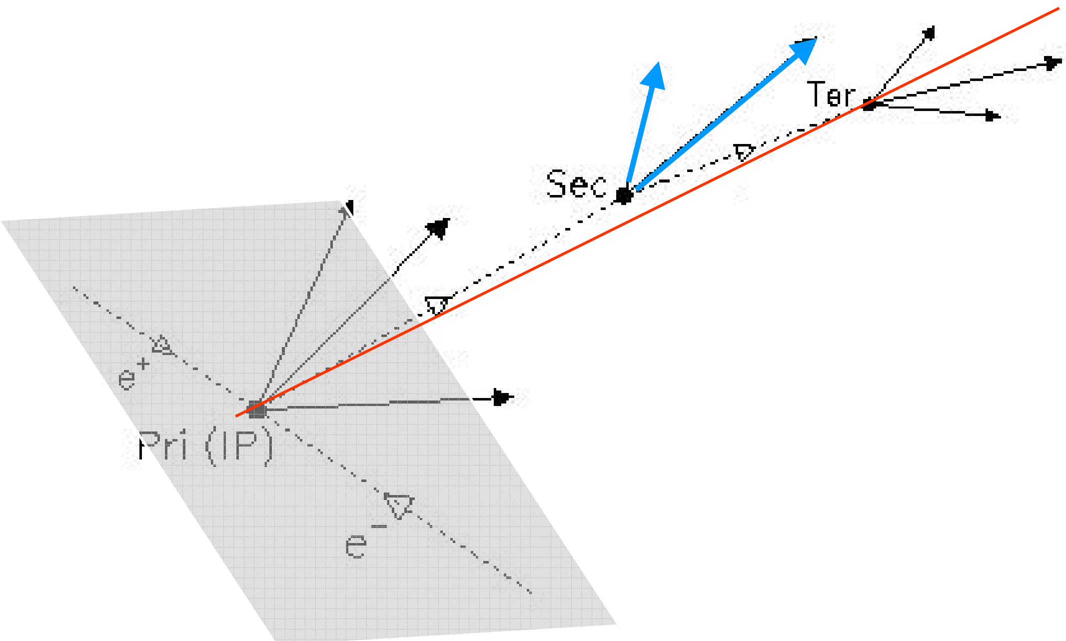
Purity of reconstructed track-vertex association (%)

MC track origin	Reconstructed track-vertex association: Two-vertex case			Three-vertex case			
	Pri.	**Sec.**	**Iso.**	**Pri.**	**Sec.**	**Ter.**	**Iso.**
Primary	90.5	1.61	27.7	97.2	4.66	2.43	46.5
B decay	7.2	48.6	35.1	1.91	74.9	9.99	24.5
D decay	2.28	49.8	37.2	0.935	20.5	87.6	29
All above	47.3	41	11.7	36.8	32.2	24.9	6.07

Diagnostic plots and tables will permit us to choose package parameters that are tuned to full MC and reconstruction and will be used in the "central reconstruction" of MC samples for ILD optimisation.

Vertex charge reconstruction

- b-jets contain a complex decay chain, from which the charge has to be found
- in the 40% of cases where b quark hadronises to charged B-hadron, quark sign can be determined by vertex charge



- need to find all stable tracks from B decay chain:
  - define seed axis
  - cut on L/D (normalised distance between IP and projection of track POCA onto seed axis)
  - tracks that form vertices other than IP are assigned regardless of their L/D

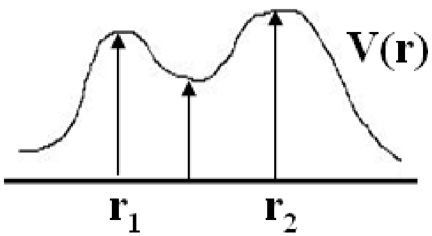
- need vertex finding as prerequisite (definition of seed axis)
- in most analyses, only calculate charge for jet of specific flavour: need flavour tagging
- probability of mis-reconstructing vertex charge is small for both charged and neutral cases

ZVTOP algorithm

- Find all maxima for $f_i(\mathbf{r})\, f_j(\mathbf{r})$ with $\boldsymbol{i}, \boldsymbol{j} = 0 \ldots N$ and collect spatial points $\mathbf{r}_{ij}$ for cases of good vertex fit, $\chi^2_{i,j} < \mathrm{X}_0$ **TwoTrackCut**
- Search in 3D for maxima in $\mathrm{V}(\mathbf{r})$ near each such $\mathbf{r}_{ij}$
- If two such maxima fail 'resolubility criterion'

$$\frac{\min\{V(\mathbf{r}) : \mathbf{r} \in \mathbf{r_1} + \alpha(\mathbf{r_2} - \mathbf{r_1}), 0 \le \alpha \le 1\}}{\min\{V(\mathbf{r_1}), V(\mathbf{r_2})\}} < R_0$$

ResolverCut



they are merged together

- Spatially resolved clusters of $\mathrm{V}(\mathbf{r})$ maxima form candidate vertices
- Track to vertex association ambiguities are decided according to largest $\mathrm{V}(\mathbf{r})$ after $\chi^2_i < \mathrm{X}_0$ trimming **TrackTrimCut**
- Vertex that includes the IP ellipsoid is called the Primary

10