Recent developments in
The LCFIVertex Package

Roberval Walsh
The University of Edinburgh
On behalf of the LCFI Collaboration

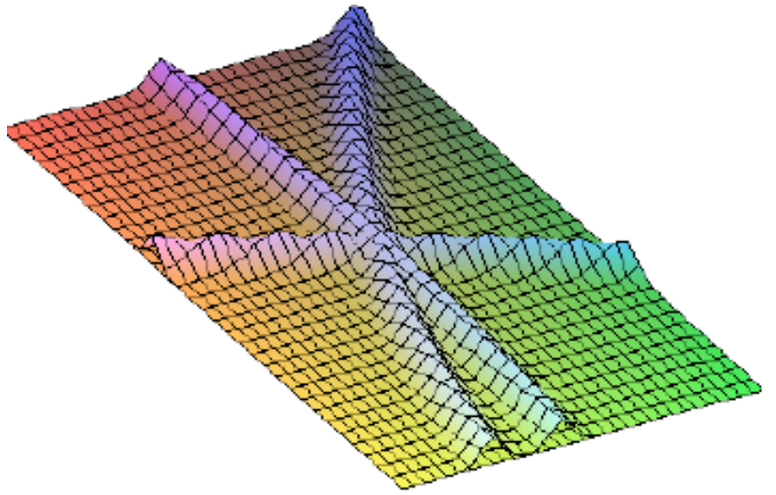
International Linear Collider
ECFA Workshop
9-12 June 2008
Warsaw, Poland

Introduction

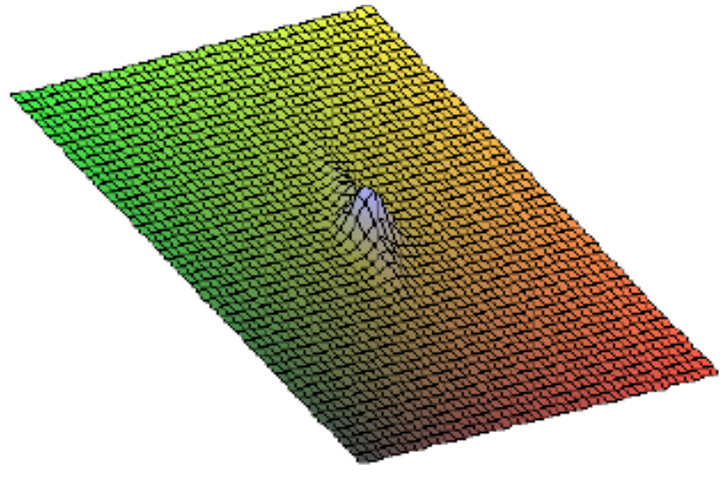
- The LCFIVertex package is a software package that uses the vertex-detector information for high-level event reconstruction at the ILC. It provides:
  - The ZVTOP vertex finder;
  - A flavour tagging algorithm based on neural networks approach;
  - Vertex charge determination for b- and c-jets.
- The code is based on Marlin and uses LCIO for input and output.
- The code and the networks are available from a CVS repository under *marlinreco* and *tagnet*, respectively: http://www-zeuthen.desy.de/lc-cgi-bin/cvsweb.cgi

ZVTOP vertex finder: ZVRES

- ZVTOP (*D. Jackson, NIM A388, 247*) consists of two branches:
  - ZVRES: General algorithm that can handle arbitrary multi-prong topologies.
    - Each track is represented as a gaussian "probability tube" in 3D where the width of the tube is the error of the track.

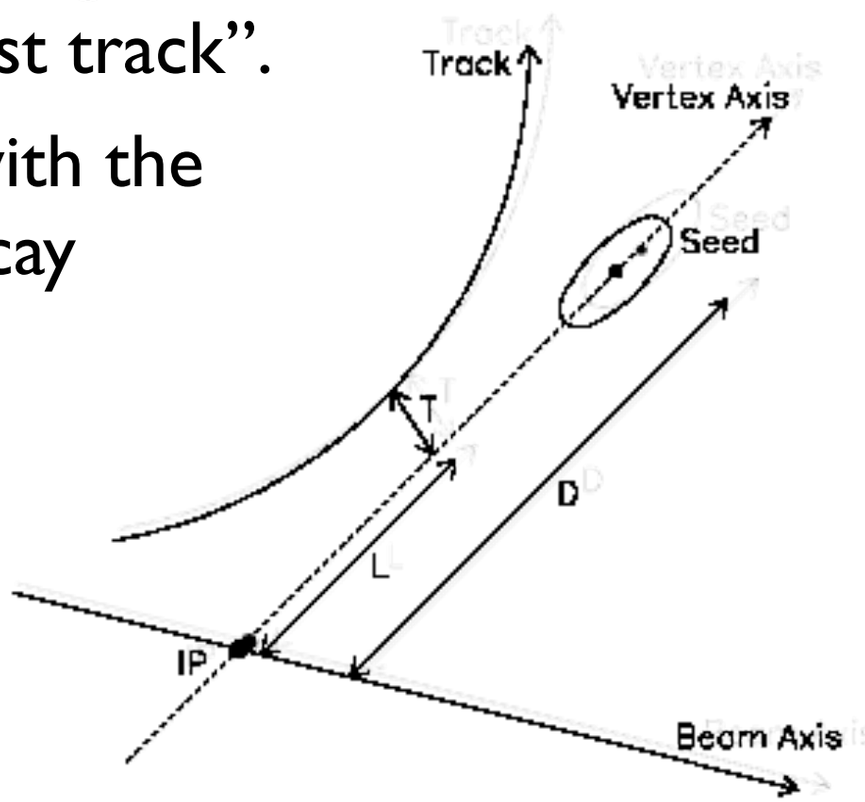


  - The track "tubes" are combined into a "vertex probability function". A maximum of this function is searched in 3D-space and the χ^2 of the vertex fit is minimised iteratively.



ZVTOP vertex finder:ZVKIN

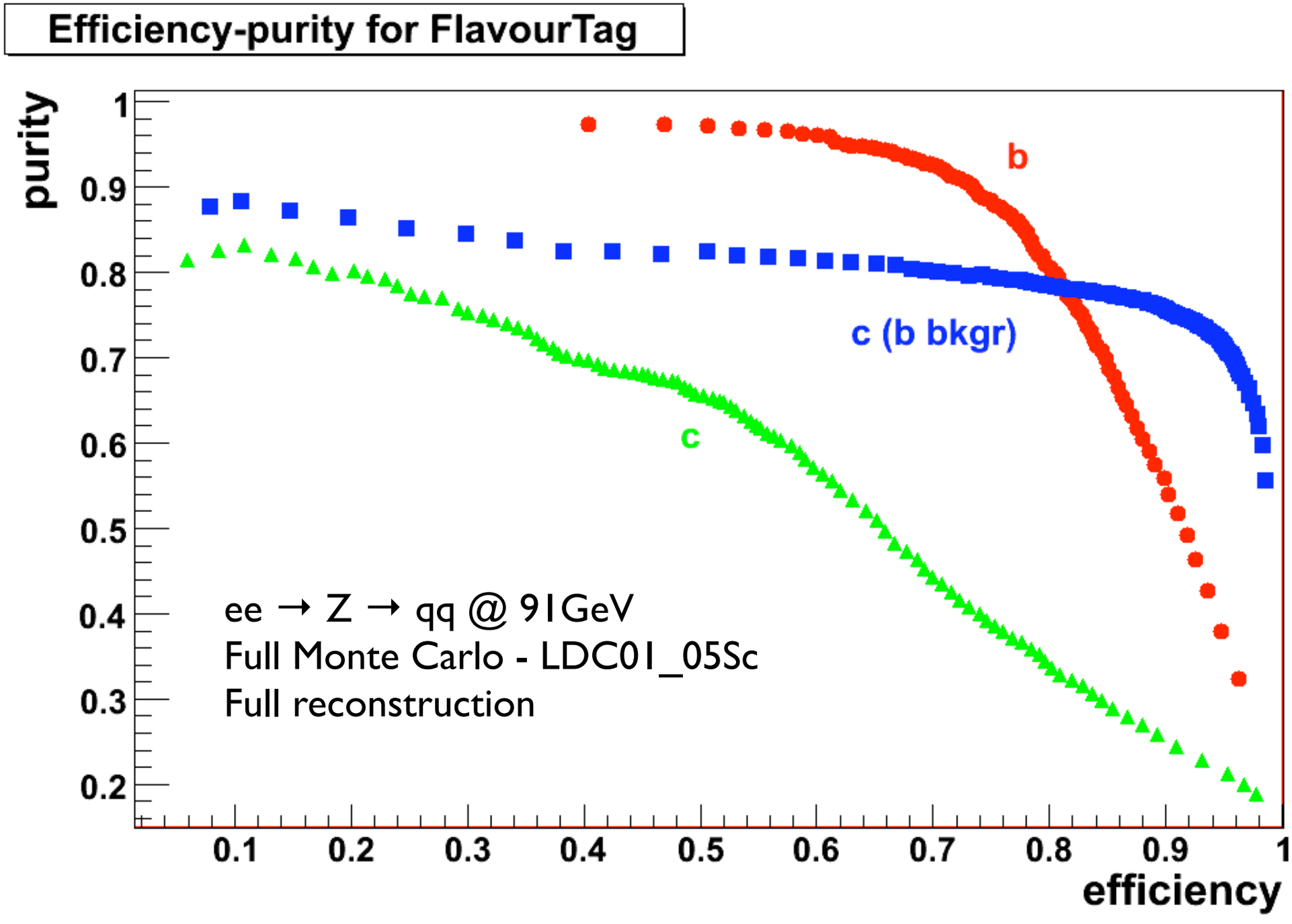
- ZVKIN (ghost track): Specialised algorithm to reconstruct vertices in b-jets with two subsequent one-prong decays. e.g. in a IP → B → D decay chain.
  - Tracks not associated to any vertex but with small values (~50μm) of the 3D transverse impact parameter T are likely to come from a B decay chain.
  - The best estimate of the direction of flight of the B-hadron is found and a finite width is assigned → "ghost track".
- The tracks in the jet are combined with the ghost track the IP to build up the decay chain along the ghost direction.
- Should improve flavour tagging efficiency, as well as the vertex charge determination.



Flavour tagging

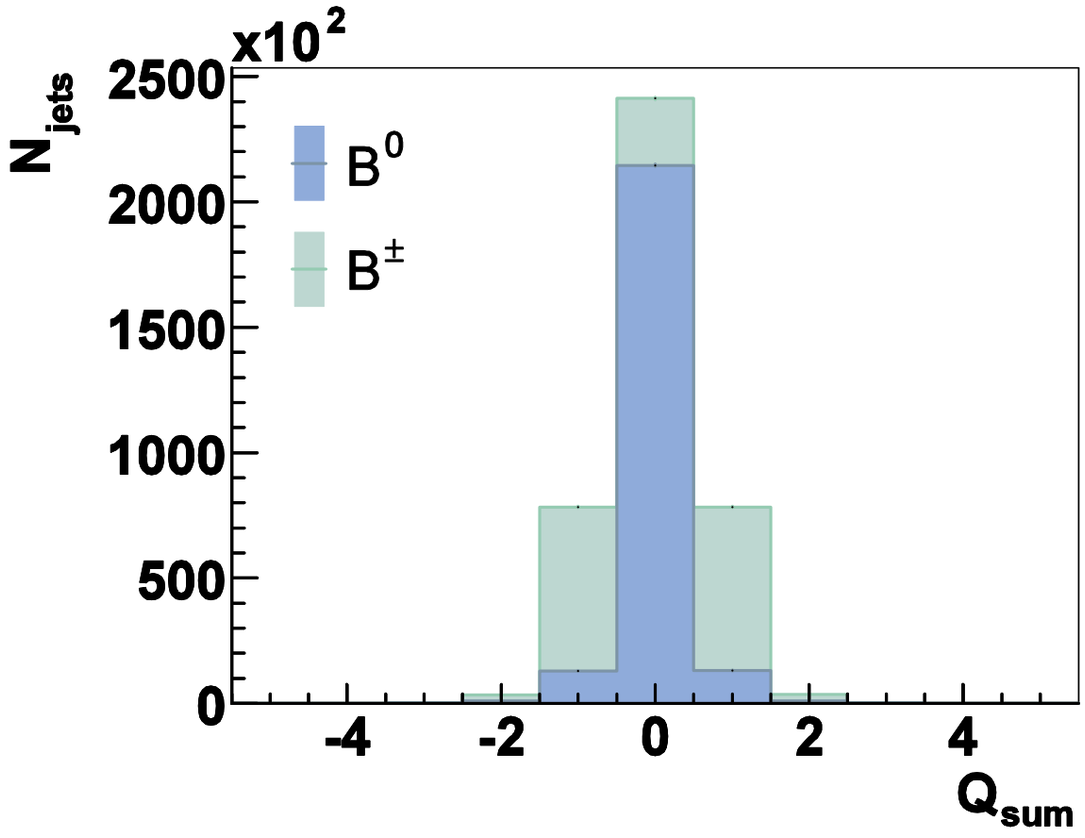
- The LCFIVertex package also incorporates a flavour tagging code (*R. Hawkings, LC-PHSM-2000-021*) using a neural network approach to discriminate between b-, c- and uds- or gluon-jets.
- Neural network input variables:
  - secondary vertex is found: p_T corrected mass, momentum, decay length and decay-length significance;
  - only IP is found: momentum and impact parameter significance in R-Φ of the two most significant tracks in the jet;
  - in all cases: joint probability in R-Φ and z (combined probability that each track of a jet comes from the primary vertex) .

Flavour tagging



Vertex charge

- The vertex charge limited to charged B-hadrons (40% of the b-jets); to reconstruct the vertex charge it is necessary to find all stable tracks from the B-decay chain.
- Probability of mis-reconstruction of the vertex charge is small for both charged and neutral hadrons.
- To extend for neutral B-hadrons, a charge dipole technique is required (under development).



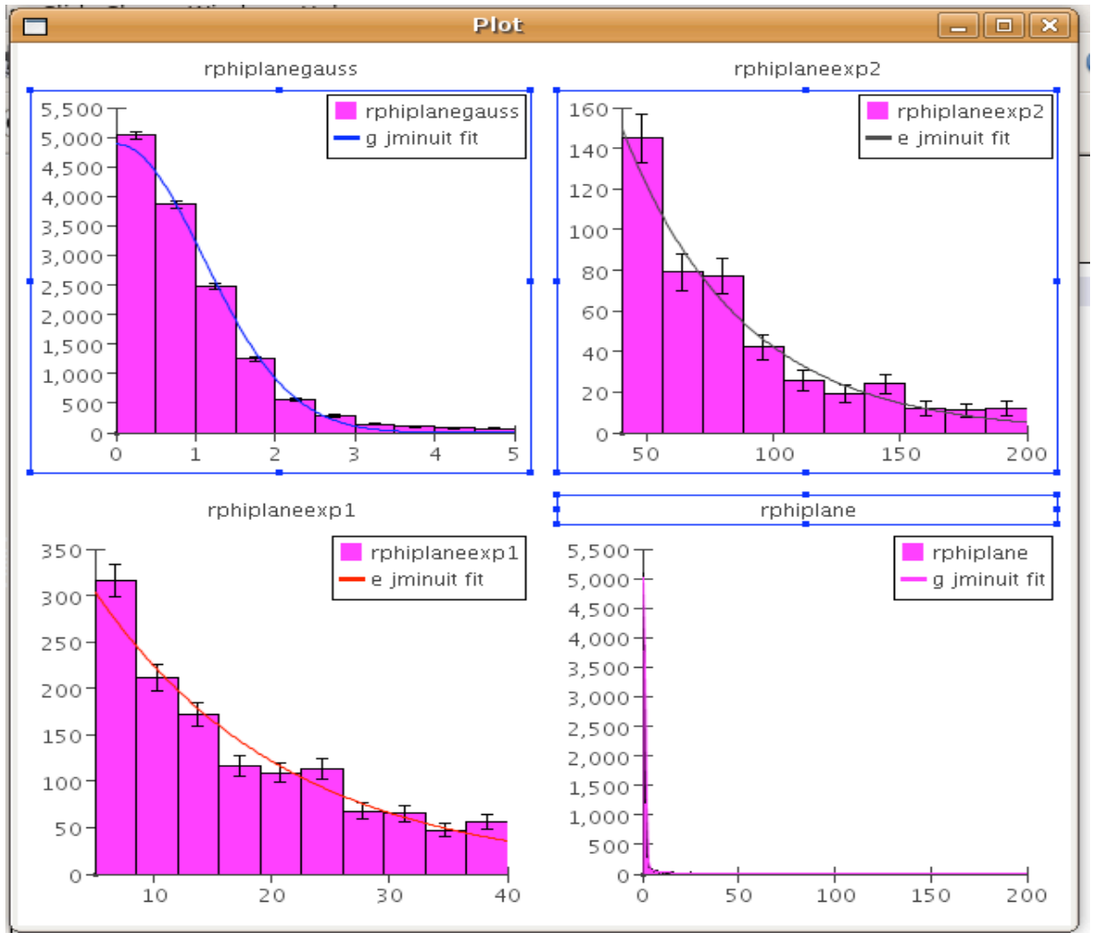
(SGV fast monte carlo sample
ee → Z → qq @ 91GeV)

Recent developments

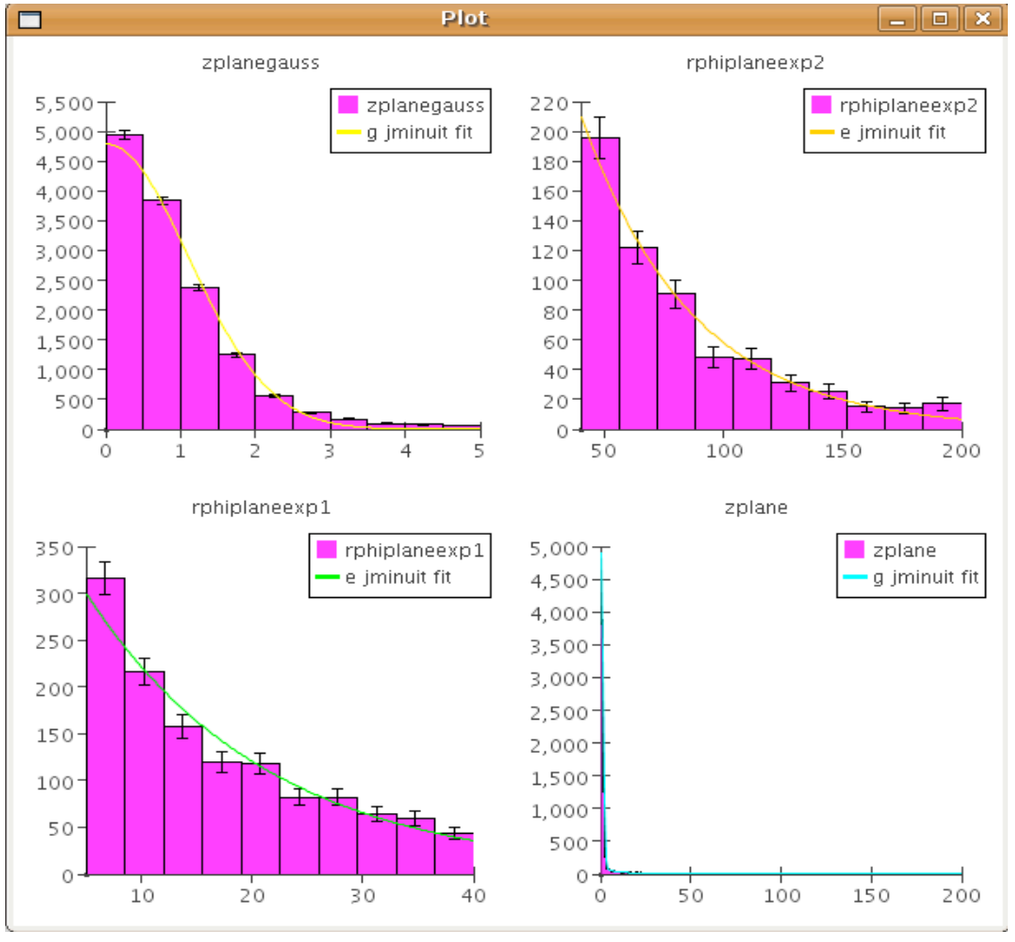
Flavour tagging: joint probabilities

- SignificanceFit processor (E. Devetak): Parameters used in the joint probability can be recalculated by fitting negative impact parameters in R-Φ and z using a gauss+exp+exp function.
- Results displayed on the screen.

R-Φ



z



Flavour tagging: NN importance

- The relative importance of the neural network inputs for a given data set is now calculated in the FlavourTag processor (E. Devetak). The results are displayed on the screen.
- The calculation of the relative importance follows the approach in the Toolkit for Multivariate Data Analysis with ROOT (TMVA) package (arXiv physics/0703039, page 65):

6.8.4 Variable ranking

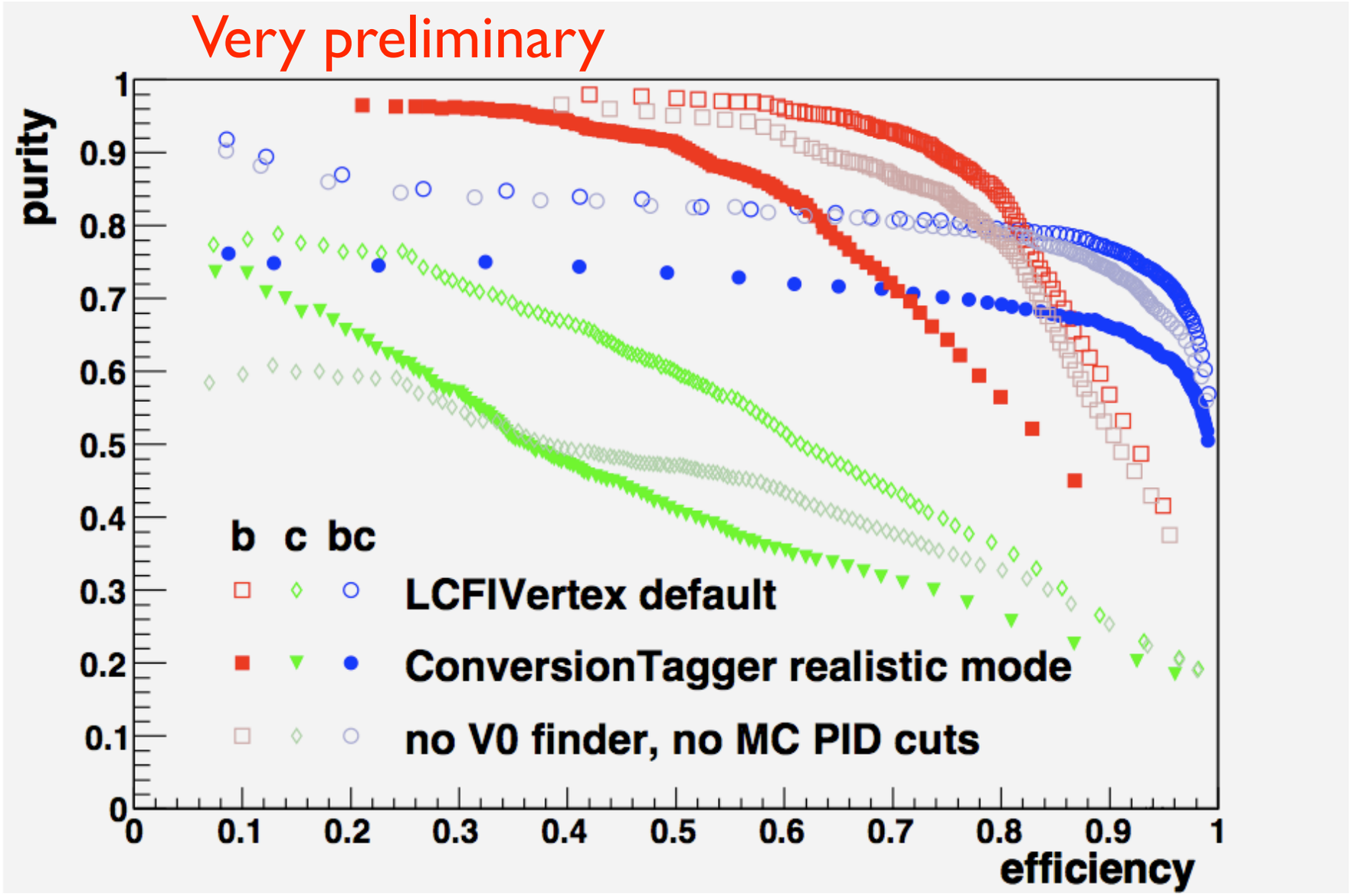
The MLP neural network implements a variable ranking based on the sum of the weights-squared of the connections that leave the variable input neuron. The importance I_i of the input variable i is given by

$$I_i = \overline{x}_i^2 \sum_{j=1}^{n_{\rm h}} \left(w_{ij}^{(1)} \right)^2 , \qquad i = 1, \ldots , n_{\rm var} , \tag{38}$$

where $\overline{x}_i$ is the sample mean of input variable i.

Conversions and V0 tagging

- ConversionTagger (K. Harder): algorithm towards a realistic photon conversion and V0 tagging.
- ConversionTagger currently uses far too loose cuts to identify tracks from conversions & V0s → seeming performance degradation not a real effect!
- Tuning of these cuts currently in progress.



Information in DST files

- DSTCollectionProcessor (C. Lynch): Minimal information from the LCFIVertex package stored in the ParticleIDs of the jets for the DST files from mass reconstruction.
    - FlavourTag: `"BTag", "CTag", "BCTag"`
    - FlavourTagInputs (only for # vertices >= 2):
        - `"JointProbRPhi", "JointProbZ",`
        - `"NumTracksInVertices",`
        - `"DecayLength", DecayLengthSignificance",`
        - `"RawMomentum",`
        - `"PTCorrectedMass",`
        - `"SecondaryVertexProbability"`
- TrueJetFlavours: `"TrueJetFlavour", TruePartonCharge, "TrueHadronCharge", "TrueJetFlavour"`

Further developments...

- SignificanceFit and LCFIAIDAPlotProcessor now compatible with RAIDA (E. Devetak, V. Martin).
- TrueJetFlavours collection: Combined MC truth with 1 LCFloatVec per jet with named components replaces separate collections (E. Devetak, V. Martin): `"TrueJetFlavour", TruePartonCharge, "TrueHadronCharge", "TrueJetFlavour"`
- Minor bug fixes.
- Current version of the LCFIVertex package is v00-02-03-dev.

LCFIVertex tuning and NN training

- Aim: Improve flavour tagging and vertex charge reconstruction.
  - How: Vary the track selection and other parameters for the primary vertex (IPFIT) and secondary vertices (ZVRES) reconstruction, and flavour tag inputs (FTI).
    - Studies in progress.
- Training of the Neural Networks:
  - Main problem is the long execution time. Grid sites have restrictions.
  - Modifications in the training code to split up the run is needed.

Summary

- The LCFIVertex package is up and running.
- Recent developments in the LCFIVertex package towards more realistic analysis.
- New codes to provide better tuning of the parameters and more information is delivered to the user.
- Minimal information for physics analysis can be stored in the DST files.
- Tuning of parameters in progress.
- Current version of the LCFIVertex package is **v00-02-03-dev**.

Extra material

Flavour tagging: joint probabilities

- The purposes of the algorithm is to determine the probability that a jet of tracks comes from the primary vertex. In a first step we determine the probability that each track of the jet comes from the primary vertex:

$$P_i = \frac{\int_{b/\sigma}^{\infty} f(x)dx}{\int_{0}^{\infty} f(x)dx}$$

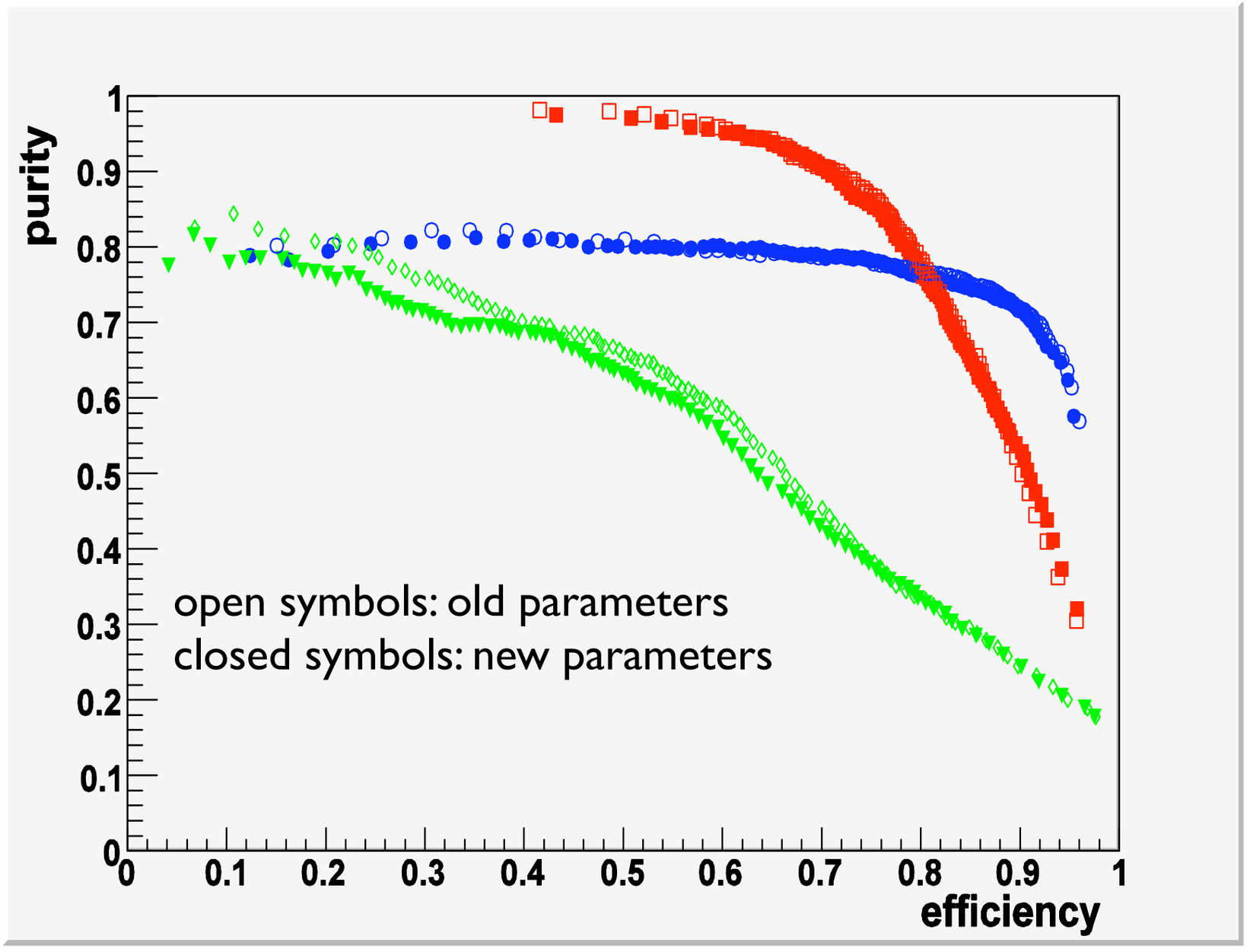
where b is the impact parameter and σ is the estimated error on b. f(x) is the distribution of impact parameters significances (modelled as a Gaussian with added exponential tails).

- Then combine the probabilities of each track into a joined jet probability for N tracks:

$$P_J = \prod_{i=1}^{N}(P_i)\sum_{m=0}^{N-1}\frac{(-\ln(\prod_{i=1}^{N}(P_i)))^m}{m!}$$

Flavour tagging: joint probabilities

- Degradation: re-train of the neural networks is needed.



LCFIVertex package tuning

(present track selection)

Description	xml parameter names	ipfit.xml	zvres.xml	fti.xml
Cut on χ^2/ndf of track fit	a1_Chi2OverDOFEnable a2_Chi2OverDOFCutLowerThan a3_Chi2OverDOFCutValue	False True 10	False True 10	False True 10
Cut on d0 (RΦ impact parameter)	b1_D0Enable b2_D0CutLowerThan b3_D0CutValue	True False 50(mm)	True False 20(mm)	True False 10(mm)
Cut on d0 error	c1_D0ErrEnable c2_D0ErrCutLowerThan c3_D0ErrCutValue	False False 0.025(mm)	True False 0.25(mm)	False False 0.025(mm)
Cut on z impact parameter	d1_Z0Enable d2_Z0CutLowerThan d3_Z0CutValue	True False 50(mm)	True False 20(mm)	True False 20(mm)
Cut on z imp parameter error	e1_Z0ErrEnable e2_Z0ErrCutLowerThan e3_Z0ErrCutValue	False False 0.025(mm)	False False 0.025(mm)	False False 0.025(mm)
Cut on p_T of the track	f1_PTEnable f2_PTCutLowerThan f3_PTCutValue	False True 0.1 (GeV)	True True 0.1 (GeV)	True True 0.1 (GeV)
Cut on K_S and Λ decay tracks	h1_MCPIDEnable h2_CutPIDS h3_MonteCarloLCRelationCollection	False --- ---	True ±310, ±3122 LDCTracksMCP	True ±310, ±3122 LDCTracksMCP