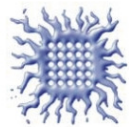
PID Tools

S. Lukić

Vinča institute of nuclear sciences, University of Belgrade

ILD analysis/software workshop 22-26. Feb. 2016



Section 1

Intro

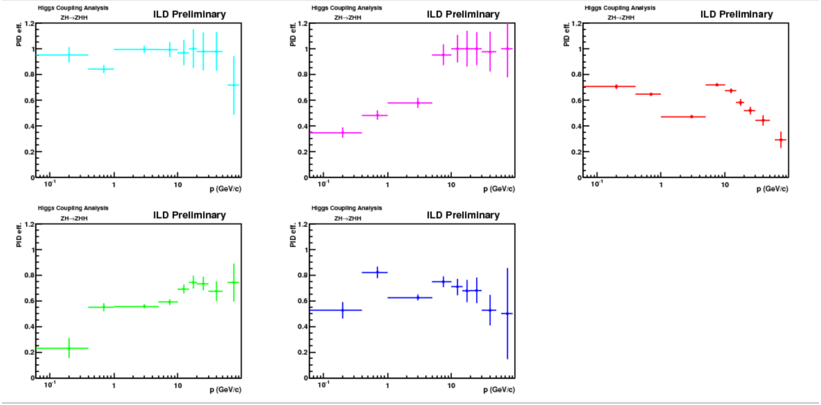
Introduction

- Main motivation: improve flavour tagging
- Set of processors and helper classes for PID
- `LikelihoodPIDProcessor` – a Bayesian likelihood classifier with 12 parameters
  - 3 "basic" parameters (total deposits from ECAL, HCAL and muon system)
  - 4 cluster-shape parameters (using shape data written by a separate processor)
  - 5 parameters based on $\mathrm{d}E/\mathrm{d}x$ in the tracker (data written by a separate processor)
- Helper classes to organise the *hypotheses* and the *data*
- MVA PID processor in development

Section 2

LikelihoodPID

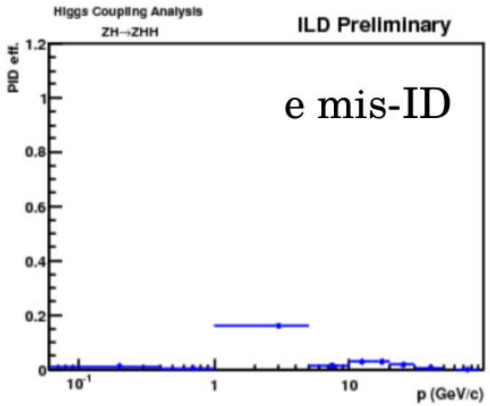
LikelihoodPID

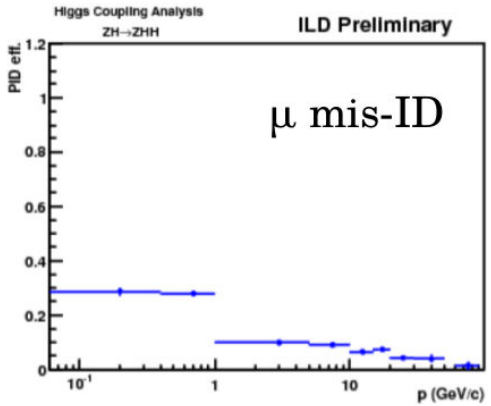


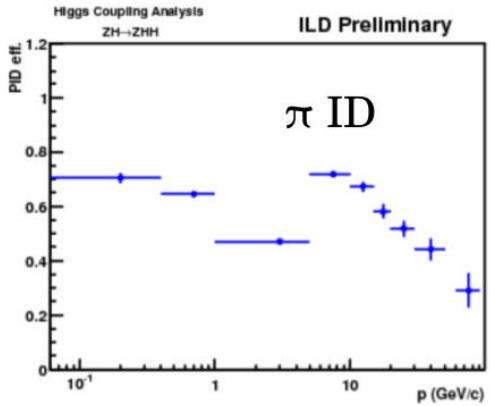
M. Kurata, feb 2016

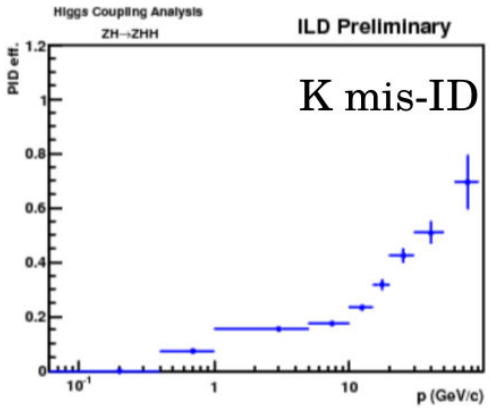
- Double Higgs at 500 GeV
- Using as much information as possible (traditional + dE/dx + Shower profile)
- Momentum dependence of PID efficiency

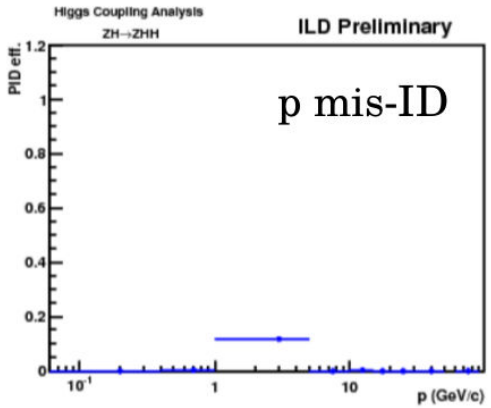
LikelihoodPID











π tracks

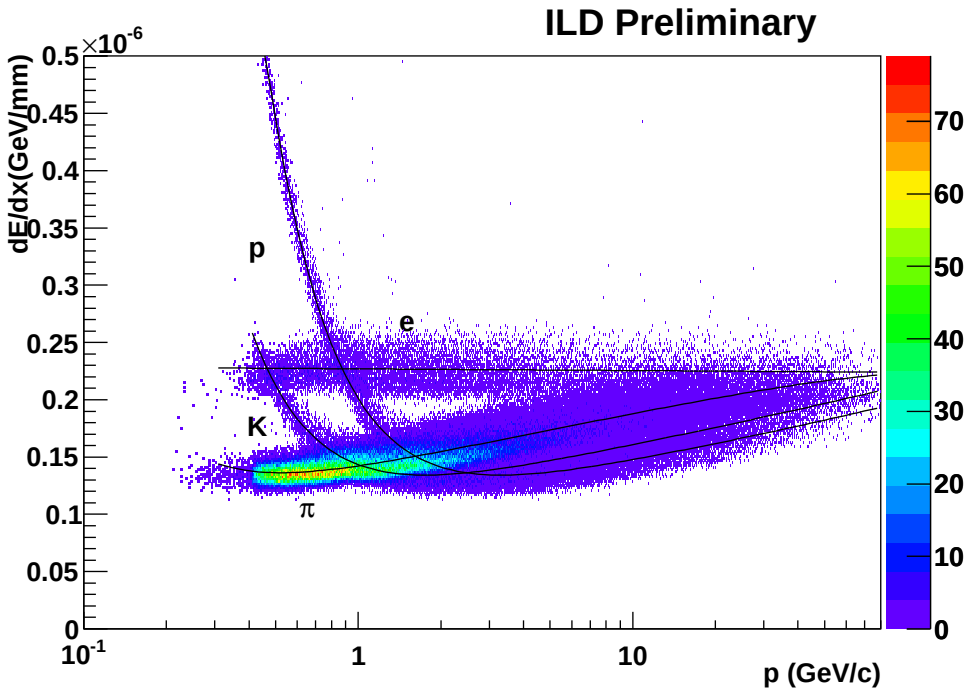
M. Kurata, feb 2016

- Double Higgs at 500 GeV
- ID and mis-ID efficiency of pion tracks

Section 3

Training in different regions of track momentum

Specific energy loss in the tracker



M. Kurata, feb 2016

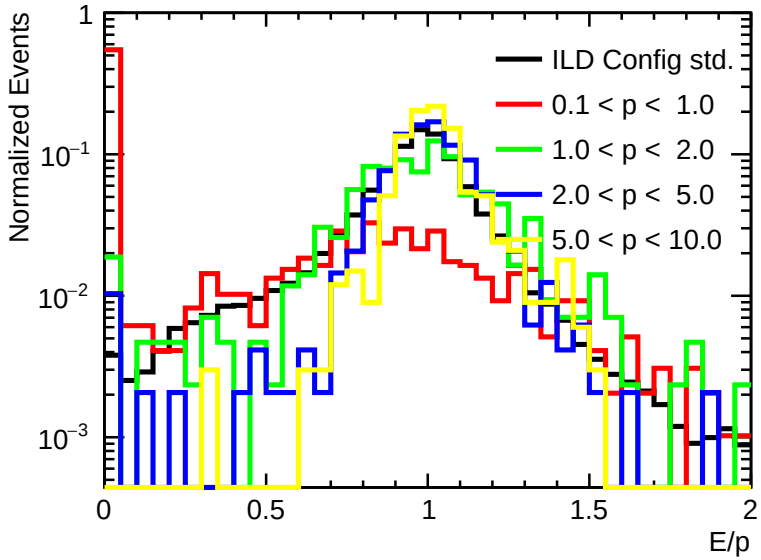
- Distinction (notably among hadrons) up to (several) 10 GeV

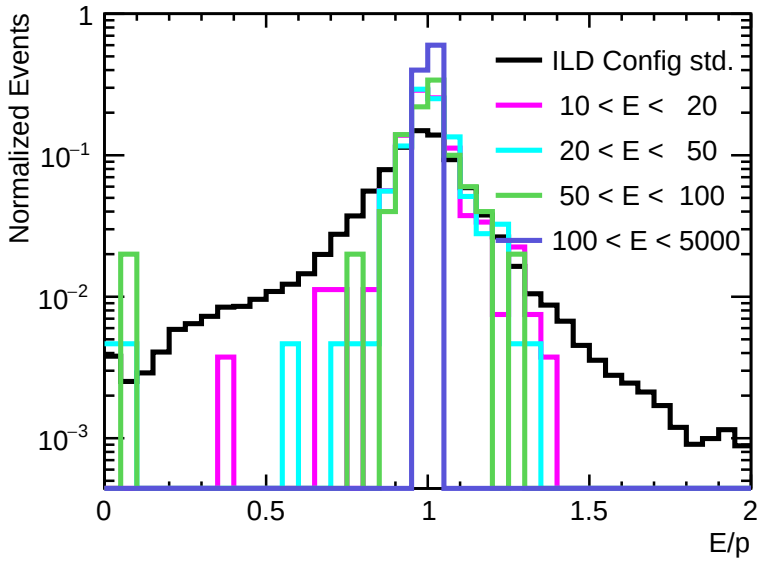
Momentum dependence of the sensitive variables

Distributions of (ECAL+HCAL)/p for ranges of p measured in the tracker

Electrons

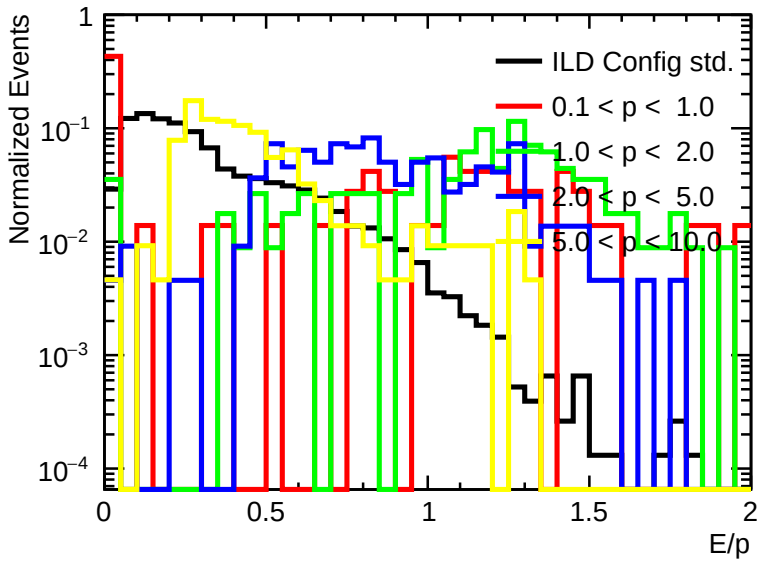
Also shown the overall distribution from ILD Standard config





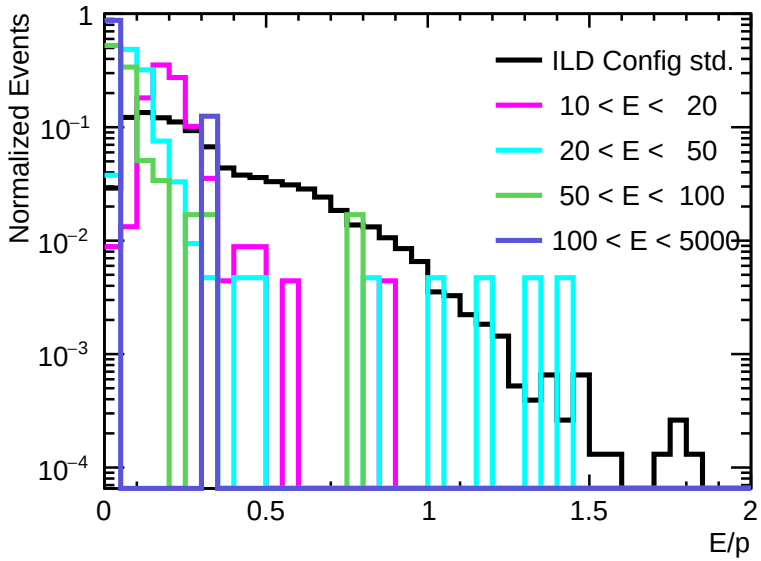
Momentum dependence of the sensitive variables

Distributions of (ECAL+HCAL)/p for ranges of p measured in the tracker



Muons

Also shown the overall distribution from ILD Standard config

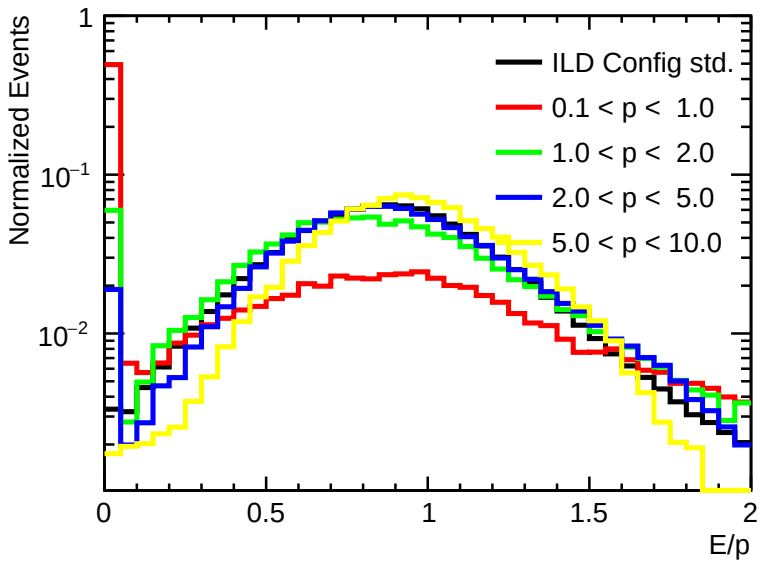


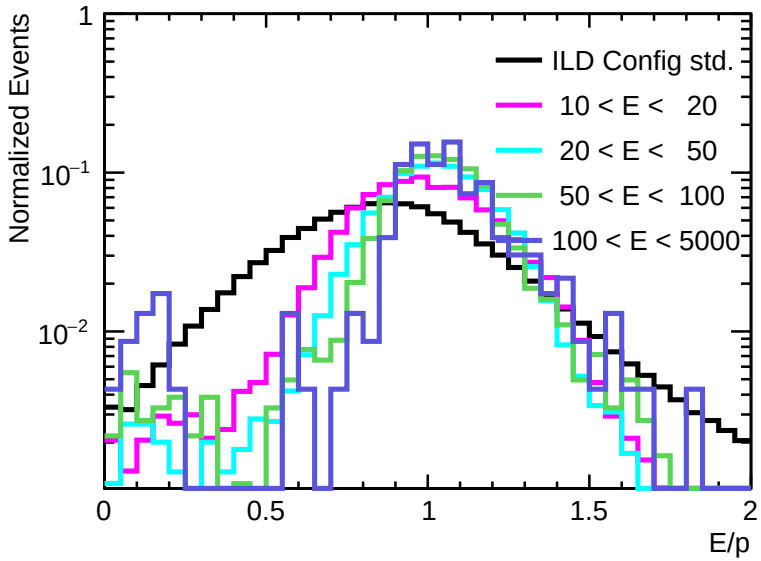
Momentum dependence of the sensitive variables

Distributions of (ECAL+HCAL)/p for ranges of p measured in the tracker

Pions

Also shown the overall distribution from ILD Standard config



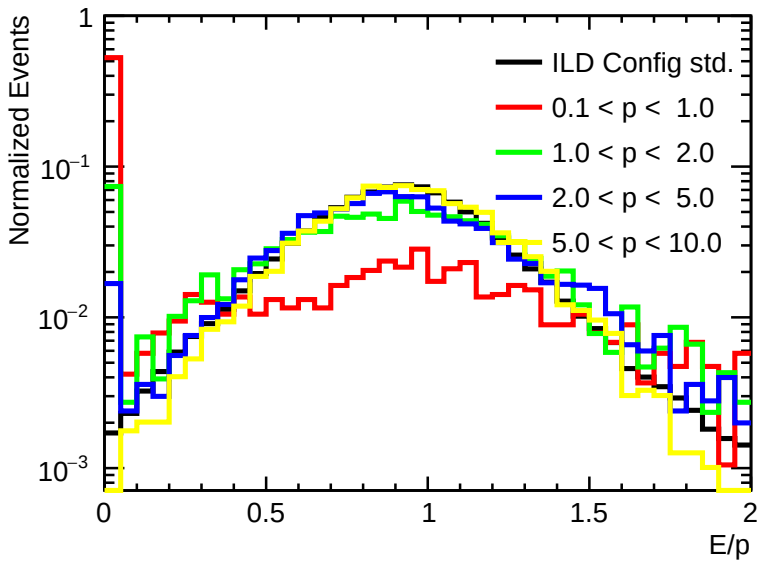


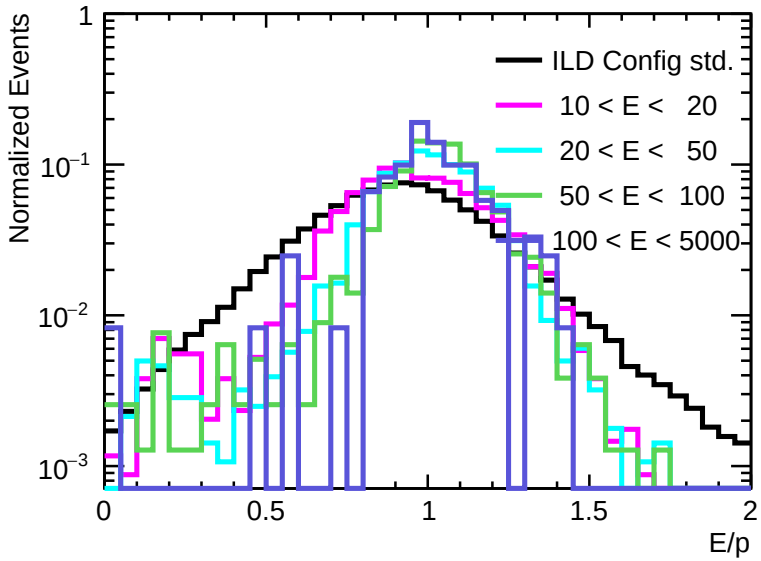
Momentum dependence of the sensitive variables

Distributions of (ECAL+HCAL)/p for ranges of p measured in the tracker

Kaons

Also shown the overall distribution from ILD Standard config



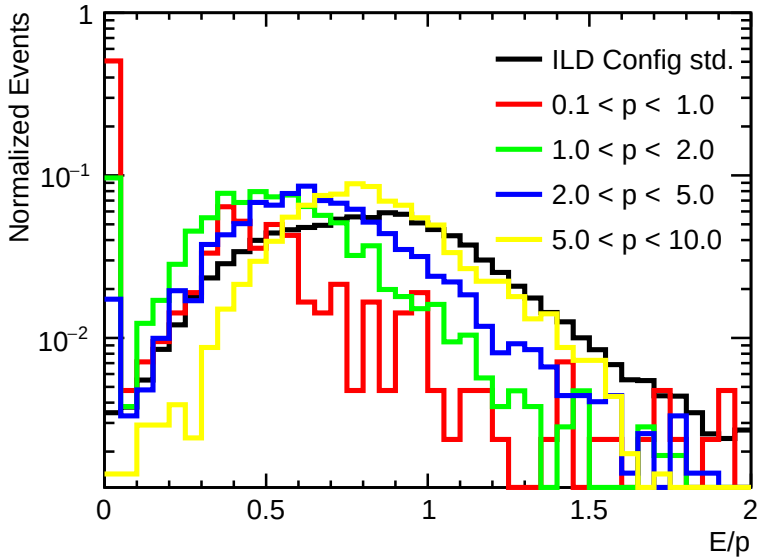


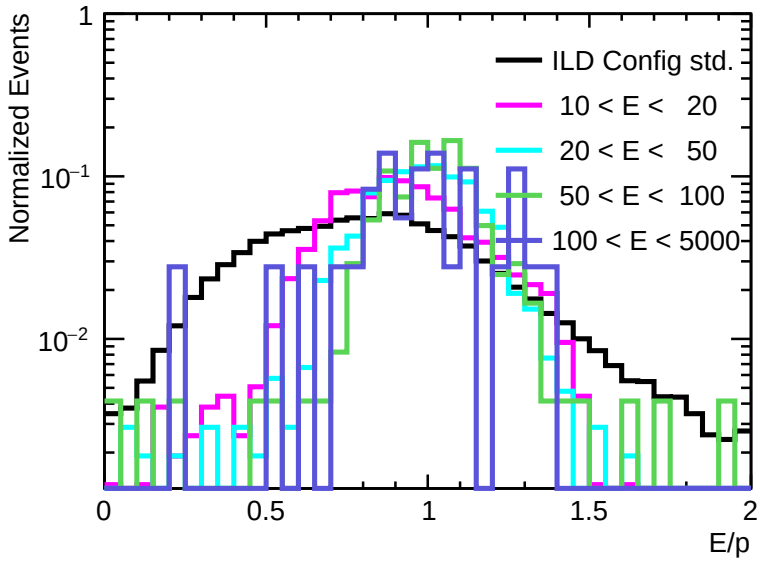
Momentum dependence of the sensitive variables

Distributions of (ECAL+HCAL)/p for ranges of p measured in the tracker

Protons

Also shown the overall distribution from ILD Standard config





Momentum dependence of the sensitive variables

- $\mathrm{d}E/\mathrm{d}x$ loses relevance at high momenta
- Calorimetric sensitive variables evolve with p – higher sensitivity should be possible if measured variables were compared to p-dependent distributions
- Hadrons more difficult to distinguish at high p
- Low-p particles may fail to reach calorimeters

Different sets of hypotheses and parameters for different intervals of measured momentum (but not **very** different!)

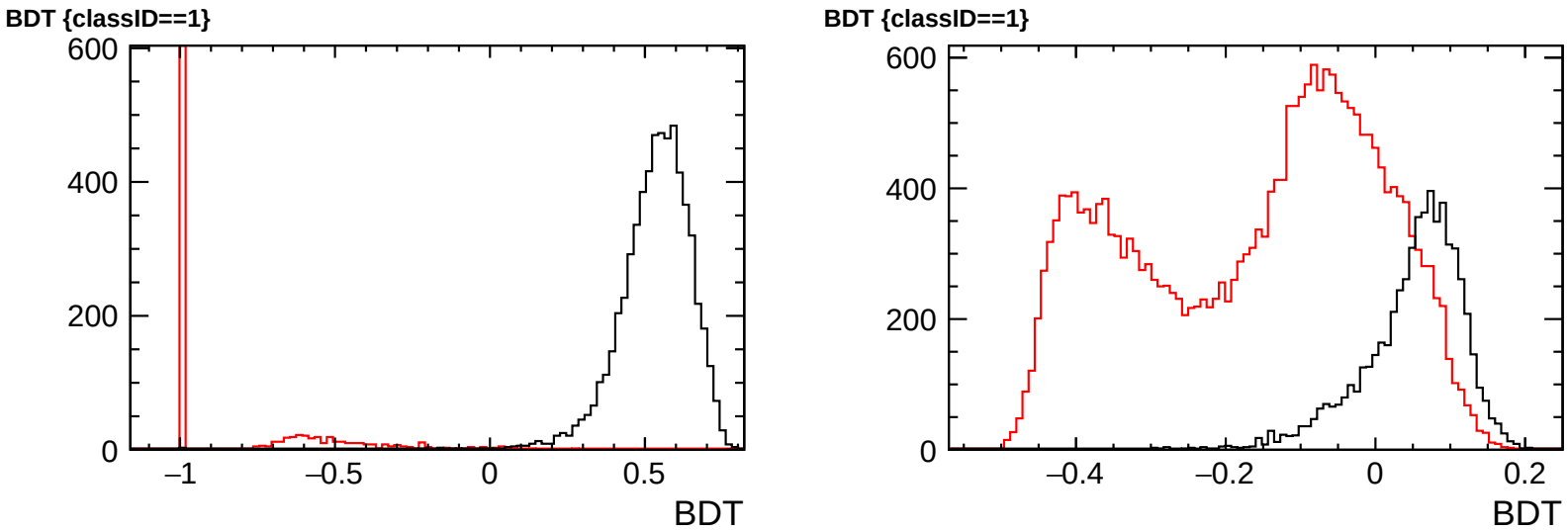
Section 4

MVA PID

MVA Strategy issues

- Multiple categories
  – different possible ways to proceed
    - Multiple trainings with 2 nodes (example in the following slide)
    - Single ANN with multiple nodes (implementation?)
- Training sample:
    - Single particles for the central analysis-independent reconstruction
    - Physics sample: analysis dependent – training by user
- Training processor available to the user

A possible classification strategy



a) b)

Signal and background BDT output for a) electrons, b) kaons in the interval $2 < p < 10\,\text{GeV}$

Only particles with associated clusters and with $\mathrm{d}E/\mathrm{d}x > 0$ used for training

- Separate training for each hypothesis vs. all others
- Write the result in the form of parameters that can be used to reconstruct probabilities of different particle-type hypotheses for the PFO, e.g.

$$\frac{\int_x^1 f_s(x')\,\mathrm{d}x'}{\int_x^1 f_s(x')\,\mathrm{d}x' + \int_x^1 f_b(x')\,\mathrm{d}x'}$$

Section 5

Summary

Summary

Status

- Likelihood PID advanced
- new processors and features under development

Outlook

- MVA PID:
  - Strategy details to be optimised
  - Minimise number of variables
  - ANN classifier with multiple response nodes? Implementation?
- Benchmark detector performance
  - Impact on flavour tagging (see talk by Masakazu in this session)
  - impact on JER using mass constraints in identified decay chains (see talk by G. Wilson on Tuesday)

Thanks!