

High-Level Reconstruction: Update since July

Jenny List
DESY
14.10.2015



ILD SW/ANA meeting

Towards a new reconstruction



Reminder: Structure agreed at Workshop

1. new standard reconstruction:
SIM -> REC, DST: digitization, full reconstruction
2. re-dsting:
REC -> DST:
add new features on DBD REC files which require HITS
dE/dx, cluster shapes, PID
this becomes possible since it is now allowed to
write out updated collections!
3. post-dsting:
DST -> postDST: collect steering examples for running
high-level reconstruction which is analysis-dependent:
isolated leptons, overlay removal, jet finding, tau finding,
pi0 finding, flavour tag

NEW – “head” installations

- in order to facilitate testing of new software, Frank kindly installs svn-head versions on DESY afs:

```
ll /afs/desy.de/project/ilcsoft/sw/[operating sys]/
```

```
drwxr-xr-x 63 voutsina af-ilc 4096 Jun 26 10:55 v01-17-07
```

```
drwxr-xr-x 64 gaede af-ilc 4096 Sep 16 15:45 v01-17-08
```

```
drwxr-xr-x 64 gaede af-ilc 4096 Oct 1 19:08 HEAD-2015-10-01
```

```
drwxr-xr-x 64 gaede af-ilc 4096 Oct 5 12:06 HEAD-2015-10-05
```

```
drwxr-xr-x 64 gaede af-ilc 4096 Oct 13 11:32 HEAD-2015-10-13
```

- “no warranty” ;-)
- but please use for testing and report problems!

New Reconstruction

- Background Overlay [optionally]
 - gammagamma->hadrons (unchanged)
 - pairs (done)
- Digitisation
 - all as is, apart from
 - VXD: 3 options for DBD, fastDBD, **challenge** (done)
- Tracking
 - all as is, apart from
 - SiTracking: 3 options DBD, **mini-vec**, FPCCD (done)
 - dEdx (done) [improve error estimate] - no news
 - V0/Kinks [fill all data members, medium term: improve!] - no news
- Garlic [optionally] (done)

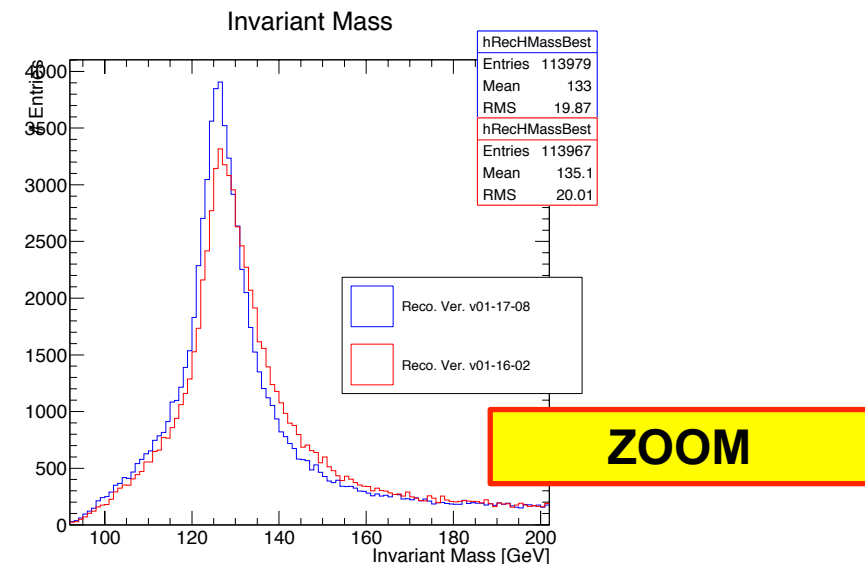
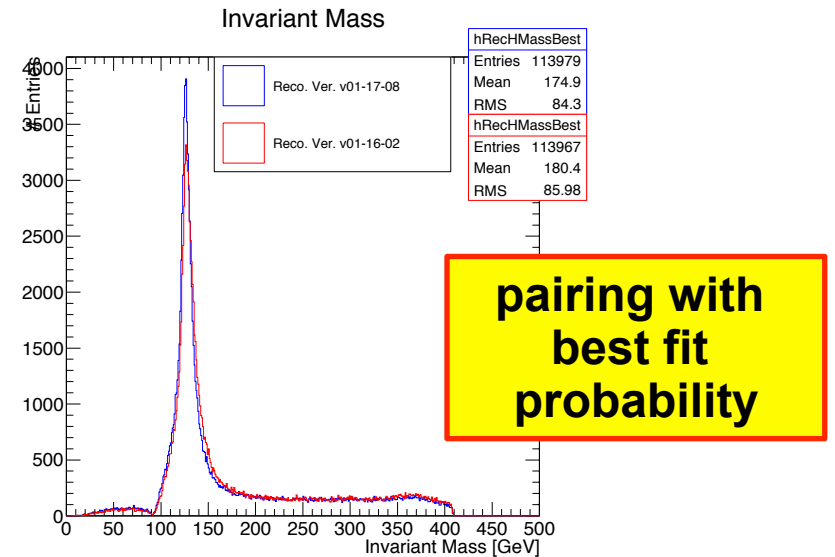
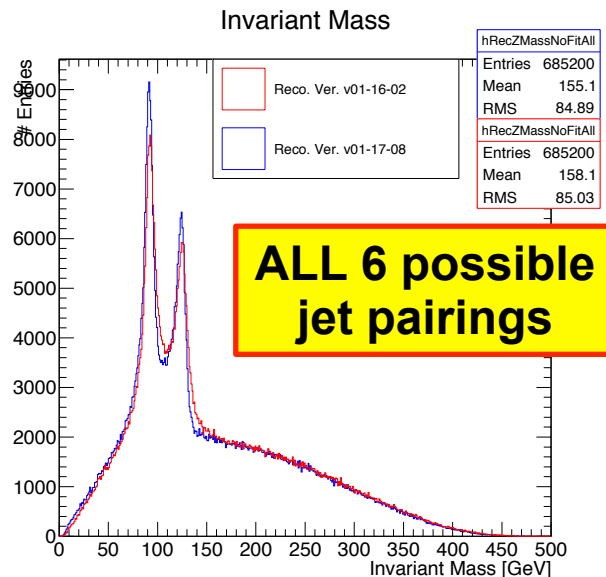
New Reconstruction (cont'd)

- Pandora
 - 3 options: new standard, **improved photons**, Garlic (done)
 - MarlinPandora/PFOCreator: fill all data members of LCIO:Clusters and ReconstructedParticles **done, but not yet in ilcsoft**
- BeamCal
 - for now as in DBD
 - new version from Andre Sailer / Andrey Sapronov: **Moritz Habermehl started to work on tuning for ILD**
 - VertexFinding
 - include adaptive vertex finding [to do] - no news
- Truth
 - RecoMCTruthLink **done!**
 - TrueJet [to be updated] - no news

Pandora with improved photons on physics

- 500 GeV, ZH→qqbb
- ILD_o1_v05, stdreco
 - **v01-16-02 (~DBD)**
 - **v01-17-08 & improved photons**
- MarlinKinfit:
 - (E,p) conservation
 - soft Z mass constraint

=> impressive improvement even in 4 jet final state!



Re-DSTing

- need to set
`<parameter name="AllowToModifyEvent" value="true" />`
- dE/dx (done) [improve error estimate] - no news
- ClusterShapes (done)
- ParticleID (done) - **needs testing -> see later**
 - “basic” (ECal/ HCal ratio etc)
 - dE/dx based
 - cluster based
 - combined
 - new: low p μ ID

Post-DST

- TauFinding
 - TaJet (taus in jet environment) (done)
 - DelphiTau (taus in low multiplicity) [to be added]
- **pi0 (eta, eta') finding (Graham)**
 - **di-photon candidate finders - done**
 - **next:**
 - **define non-overlapping subset**
 - **write improved PFO collection with di-photons replaced by mass-constrained meson 4-momentum**
- Isolated Lepton Finding (done)
- Jet Finding
- Flavour Tag

Content of Clusters and ReconstructedParticles

Status PFOCreator

- PFOCreator in MarlinPandora [Bono]:
 - in principle updated, cf presentation by Bono Sep 30
 - yesterday:
published in feature branch of Pandora git repository
<http://github.com/PandoraPFA/MarlinPandora/tree/pfoAndTrackCreatorUpdate>
 - however this not work together with ilc_install,
thus inclusion in ilcsoft installation requires merging into
master branch - time line?

Status Cluster Uncertainties

- ClusterShape class [Mikael]:
 - calculation of uncertainties / covariances on cluster main axes from hit positions turns out to be mathematically much more involved than thought at the workshop ☹
 - **math done**
 - **checks & test done**
 - **now being implemented in ClusterShape class...**
 - expect detailed presentation soon!

Integration into stdreco

- philosophy in the past:
MarlinPandora writes “the” PFO collection, later updates only in same job or by copying into new collection
 - since HLRecoWS:
update mechanism for existing collections
- ⇒ enables more easy post-processing of PFOs & clusters
- ⇒ need to set only Pandora specific information on PFOs/ clusters in PFOCreator (eg calibrated energies)
- ⇒ other information eg cluster direction uncertainties go into separate Marlin processor, as already the case for ChargedPFO4MomentumCovarianceMatrix

ILD Performance



ILD Performance Package

- Prototype by Yorgos, cf. presentation in Wednesday meeting
 - add recipe to obtain standard performance plots
 - more details than the hand-full of plots in DBD
 - **for software validation:
the collection of standard tests!**
 - for performance comparison
 - laid out detailed plan at HLRecoWS for benchmarks based on
 - physics event samples (eg H / Z / W masses, flavour tag)
 - single particle samples (e, mu, pi, K, ...)
- => not much happened so far since workshop?
(to my knowledge at least...?)**

Some first look into testing PIDTools

ILDPerformance/PID

=====

particle ID part of ILD performance

J.List, Sept/Oct 2015

needs:

- a physics sample (best with jets) to test (DST)
- insert path in scripts/PIDTree.xml
- if needed adjust collection names and/or name of output root tree
- matching GEAR file (scripts/GearOutput.xml is ILD_o1_v05)

=====

cd scripts

\$MARLIN/bin/Marlin PIDTree.xml [writes ../Results/analysis_output.root"]

cd ../macros

root -l

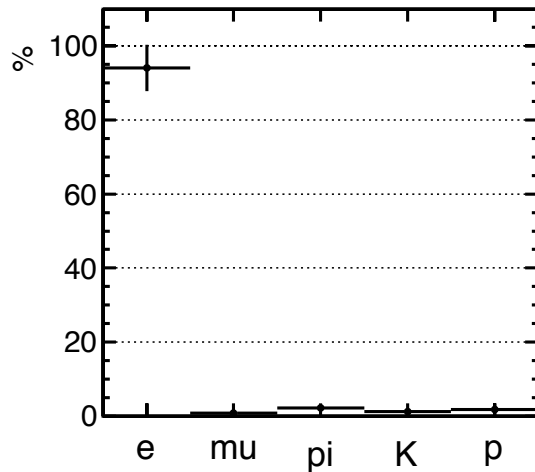
. x plotPIDs.C("../Results/analysis_output.root");

[plots PID efficiencies for the four algorithms, writes ../Results/PIDs_pdg[i]_cut[j][eps,ps,pdf].

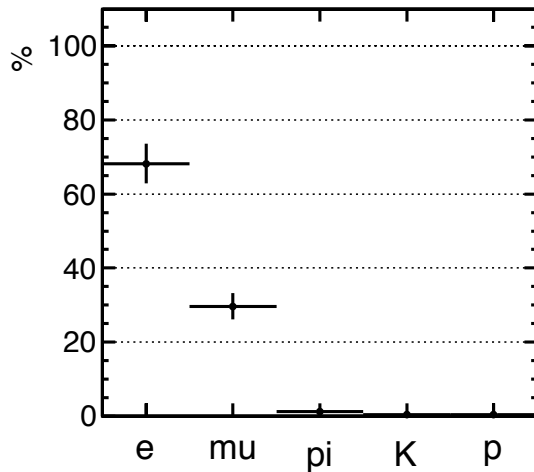
cuts on true particle can be adjusted]

Example: mumuH @ 500 GeV

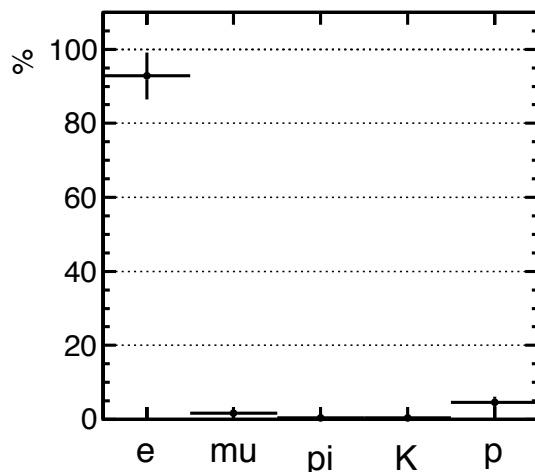
BasicPID, electrons with $P > 1$ GeV Entries 236



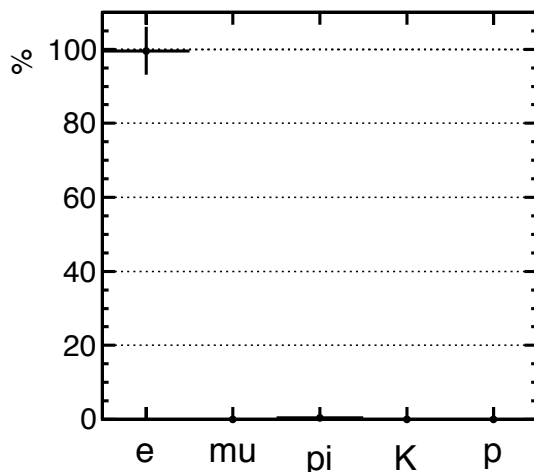
dEdxPID, electrons with $P > 1$ GeV Entries 236



ShowerPID, electrons with $P > 1$ GeV Entries 236



LikeliPID, electrons with $P > 1$ GeV Entries 236



selection:

- true **electrons** with $p > 1$ GeV
- reconstructed
- weight of RecoTruth relation $w > 0.5$

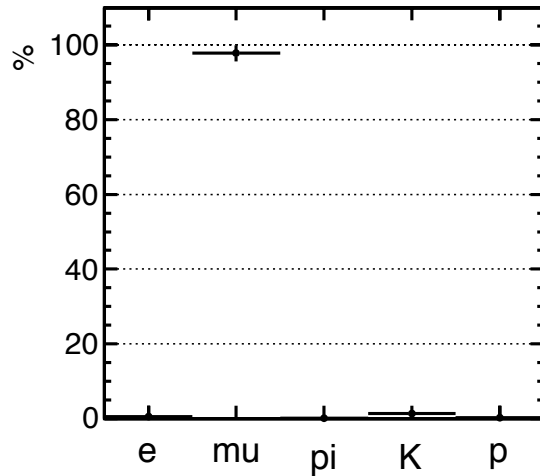
PIDs:

- from Masakazu's PIDTools
- basic: e/p, Ecal/Hcal
- dE/dx
- Shower shapes
- combined likelihood, "trained" on ZHH- \rightarrow ZbbWW*

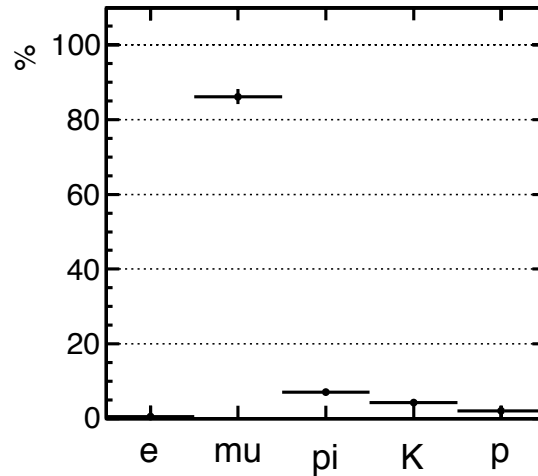
**electrons look very reasonable
combined likelihood
extremely pure**

Example: $m_{\mu\mu}H$ @ 500 GeV

BasicPID, muons with $P > 1$ GeV Entries 2054



dEdxPID, muons with $P > 1$ GeV Entries 2054



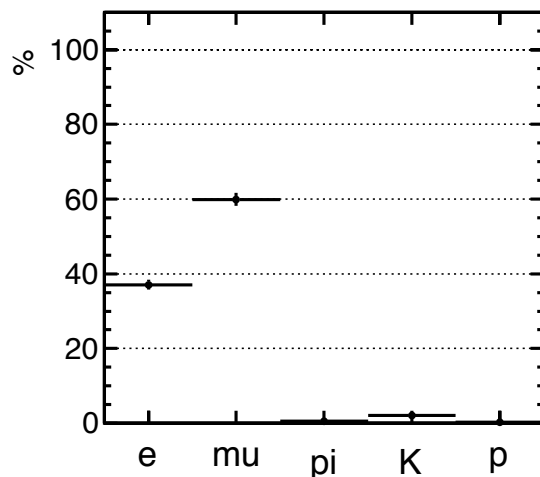
selection:

- true **muons** with $p > 1$ GeV
- reconstructed
- weight of RecoTruth relation $w > 0.5$

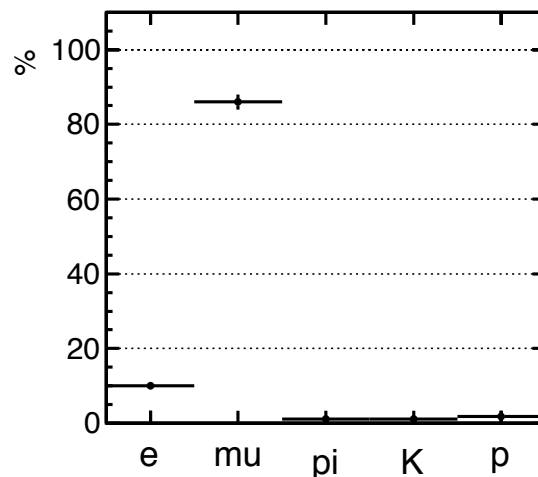
PIDs:

- from Masakazu's PIDTools
- basic: e/p, Ecal/Hcal
- dE/dx
- Shower shapes
- combined likelihood, "trained" on $ZHH \rightarrow ZbbWW^*$

ShowerPID, muons with $P > 1$ GeV Entries 2054



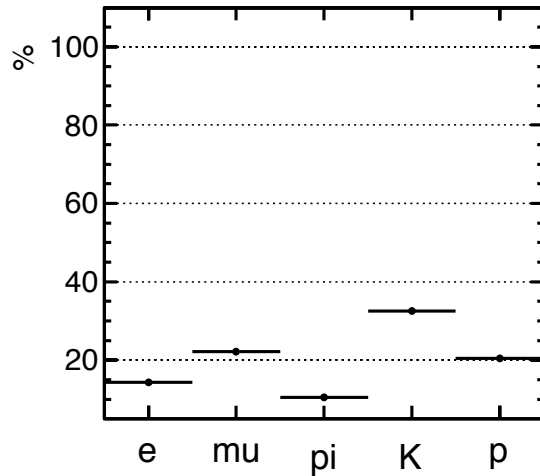
LikeliPID, muons with $P > 1$ GeV Entries 2054



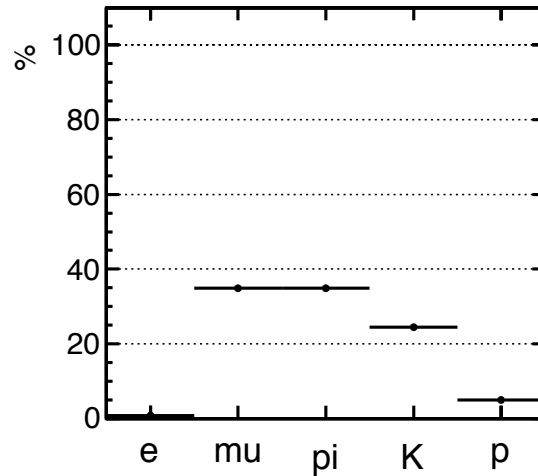
muons identified best by basic ID?!
dE/dx looks too good wrt mu/pi separation?

Example: mumuH @ 500 GeV

BasicPID, pi+- with P > 1 GeV Entries 6085



dEdxPID, pi+- with P > 1 GeV Entries 6085



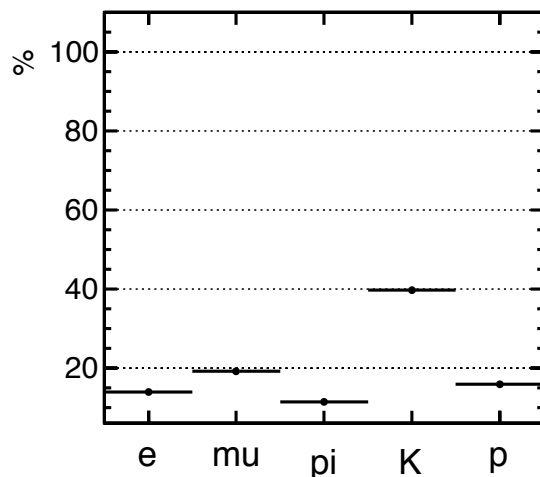
selection:

- true **pions** with $p > 1$ GeV
- reconstructed
- weight of RecoTruth relation $w > 0.5$

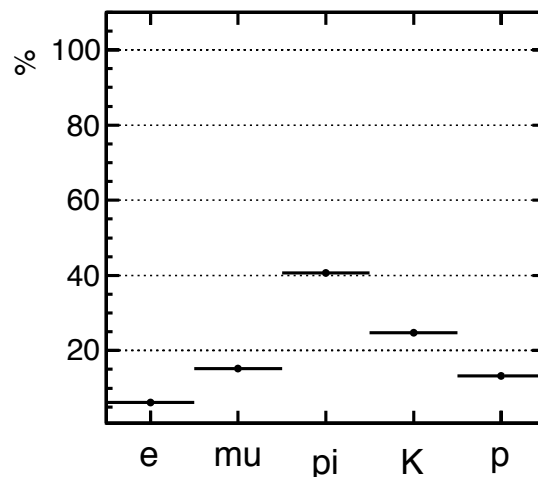
PIDs:

- from Masakazu's PIDTools
- basic: e/p, Ecal/Hcal
- dE/dx
- Shower shapes
- combined likelihood, "trained" on ZHH->ZbbWW*

ShowerPID, pi+- with P > 1 GeV Entries 6085



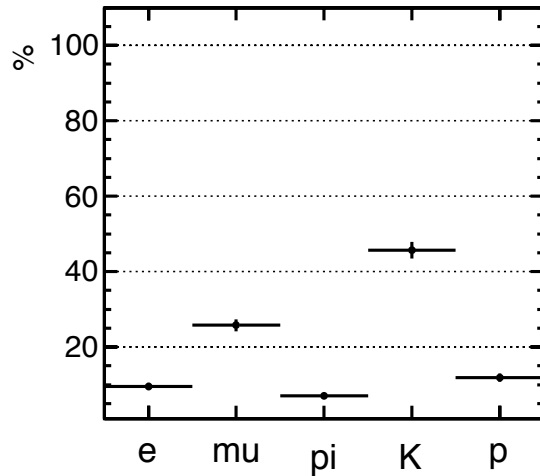
LikeliPID, pi+- with P > 1 GeV Entries 6085



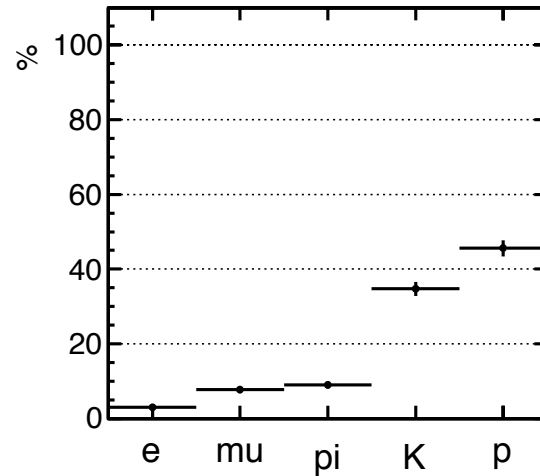
how "can" basic / shower shape identify different hadrons?
=> should only distinguish between em / mu / hadron ?

Example: $m_{\mu\mu}H$ @ 500 GeV

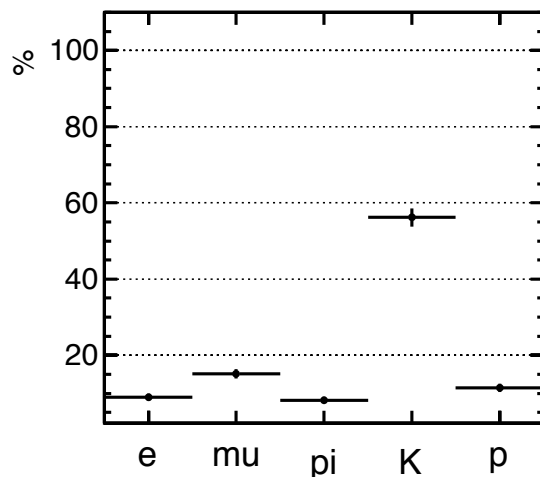
BasicPID, K^{\pm} with $P > 1$ GeV Entries 974



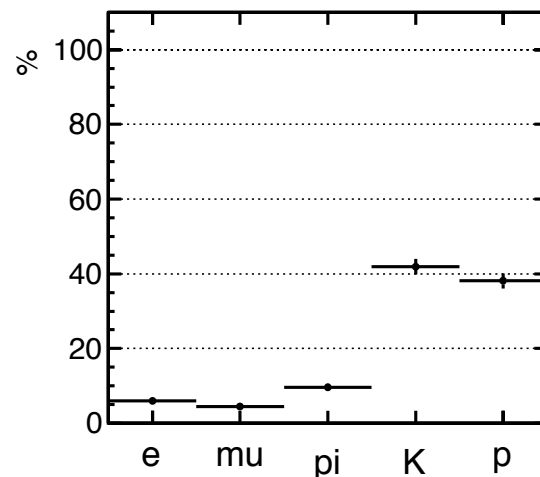
dEdxPID, K^{\pm} with $P > 1$ GeV Entries 974



ShowerPID, K^{\pm} with $P > 1$ GeV Entries 974



LikeliPID, K^{\pm} with $P > 1$ GeV Entries 974



selection:

- true **Kaons** with $p > 1$ GeV
- reconstructed
- weight of RecoTruth relation $w > 0.5$

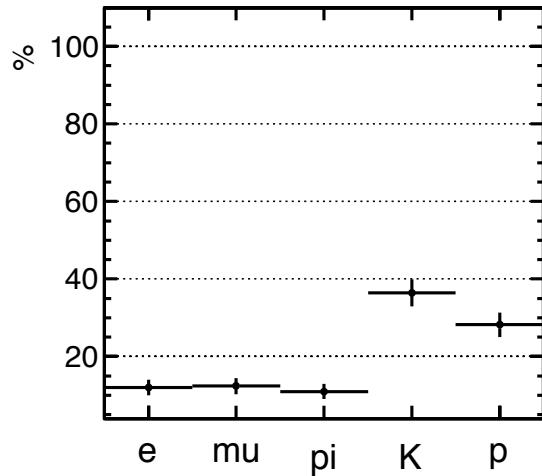
PIDs:

- from Masakazu's PIDTools
- basic: e/p, Ecal/Hcal
- dE/dx
- Shower shapes
- combined likelihood, "trained" on $ZHH \rightarrow ZbbWW^*$

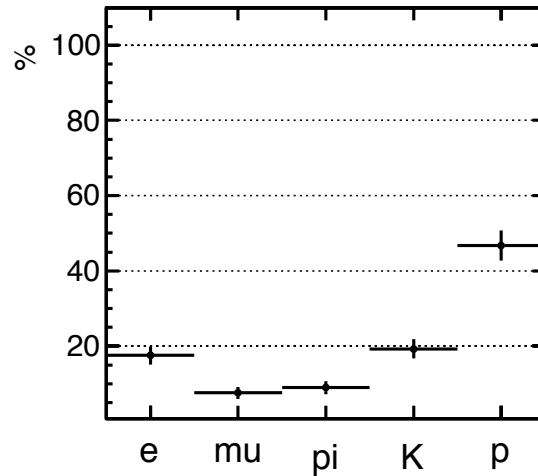
good performance of shower shape seems "random", since pions are also called Kaon most of the time....

Example: mumuH @ 500 GeV

BasicPID, protons with $P > 1$ GeV Entries 291



dEdxPID, protons with $P > 1$ GeV Entries 291



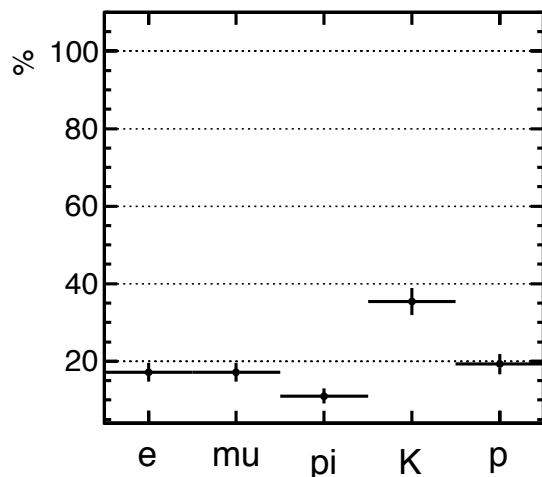
selection:

- true **protons** with $p > 1$ GeV
- reconstructed
- weight of RecoTruth relation $w > 0.5$

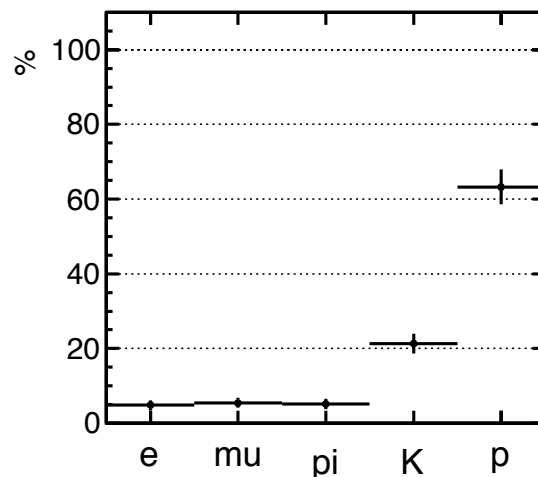
PIDs:

- from Masakazu's PIDTools
- basic: e/p, Ecal/Hcal
- dE/dx
- Shower shapes
- combined likelihood, "trained" on ZHH- \rightarrow ZbbWW*

ShowerPID, protons with $P > 1$ GeV Entries 291



LikeliPID, protons with $P > 1$ GeV Entries 291



**dE/dx looks nice,
likelihood better by
suppressing electron
misID**

Conclusion PIDs

Technical issue:

each PID currently writes the extra parameters of ALL PID algorithms – i.e. showershape writes “dEdx likelihoods” etc

⇒ very confusing (and waste of disc space...)

PIDs need another iteration

- better understand algorithm performance
- clean-up technical issues

Conclusions

- we are on a good way:
 - **saw already many improvements since HLRecoWS**
 - more things being worked on, expected soon
- however:
 - still several open issues
 - **much more testing needed!!!**
- **no sufficient effort yet: ILCPerformance**

Backup



DBD Reconstruction - Overview

```
<marlin>
  <execute>
<!-- ===== overlay gamma gamma background ===== -->
  <processor name="BgOverlay" />
<!-- ===== track digitization and tracking ===== -->
.....
<!-- ===== the new C++ tracking ===== -->
<processor name="MyClupatraProcessor" />
<processor name="MySiliconTracking_MarlinTrk" />
<processor name="MyForwardTracking"/>
<processor name="MyTrackSubsetProcessor" />
<processor name="MyFullLDCTracking_MarlinTrk"/>

<!-- ===== the post tracking patrec ===== -->
  <processor name="MyV0Finder"/>
  <processor name="MyKinkFinder"/>
.....
```

Fix average number

hadron overlay

Minivector VTX tracking

TPC, Si, Fwd tracking, combined track fit in FullLDCTracking

????? Status?

V0 & Kink finding: input to Pandora

DBD Reconstruction - Overview

```
<!-- ===== calorimeter digitization and PFA ===== >
....
<processor name="MyMainPandora"/>
<processor name="MyBeamCalReco"/>

<!-- ===== particle ID ===== -->
<!--processor name="MyPFOID" / ->

<!-- ===== full and DST output ===== >
<processor name="MyRecoMCTruthLinker"/>

<!-- ===== vertex finder ===== >
<processor name="VertexFinder"/>

<processor name="MyLCIOOutputProcessor"/>
<processor name="DSTOutput"/>
</execute>
```

PFO covariance matrix

Update Pandora & Calib Garlic?

PFO creation: PandoraPFA New BeamCal (pair bkg)

New BeamCalReco

PARTICLE ID!

Updated version, truth vertices

Link PFOs with MCTruth

track recovery, updates

VertexFinder from LCFIPlus

are we happy with DBD DST format?

REC & DST output

Content of EVENT::Cluster - TODO

getType:

should be bits from calos contributing energy

not used now -> do we need it ? – **to leave unused**

getEnergy: Pandora improved energies - **todo**

getEnergyError:

if pdg != 22/11: 60%/ sqrt(getEnergy) +3%

if pdg = 11/22: 17%/sqrt(getEnergy) + 1% ,
as used in Pandora's track-cluster matching - **todo**

get SubdetectorEnergies:

raw hit sums

split between barrel / endcaps -> todo

getHitContributions = if hit belongs to several clusters! - not used
by Pandora

Content of EVENT::Cluster - TODO

getPosition: center-of-gravity as default - ok

for photons: via cluster shape (Graham & John todo:
verify implementation in Pandora and transfer information out to
LCIO for Cluster)

getITheta/IPhi: direction of cluster main axis

getPositionError, getITheta/IPhiError: rms of cog/main axis, to
be calculated in the same place: ClusterShapes.cc -
TODO (->Mikael)

all properties will be set in MarlinPandora/.../PfoCreator.cc

routines for actual calculations:

- > eventually to MarlinUtil/ClusterShapes
- > for development: MarlinReco/Analysis/

Content of EVENT::ReconstructedParticle - TODO

currently filled in PFOCreator.cc

getType: particle "ID" by Pandora

isCompound: revise logic

- add "is not used in compound particle" = isConstituent
- todo!

momentum / energy: from track or cluster depending on charge

getMass: set independently!

getCharge: as is

getCovariance:

charged PFOs: implemented by Tino – todo: put in MarlinUtil/ (MarlinReco/ Analysis) and use in PfoCreator.cc - Tino

neutral PFOs: from cluster uncertainties – todo

getReferencePoint (todo?):

charged PFOs: z0 and (x0,y0) from (d0, phi0)

neutral PFOs: cluster position (cog or improved from shower shape)

Content of EVENT::ReconstructedParticle - TODO

getParticleIDs: as discussed

getParticles: if compound...

getTracks, getClusters: ...

getStartVertex, getEndVertex (todo):

- should be filled by Pandora for V0s, Kinks etc

- should be filled by vertexing for the rest

 - > needs to be able to update PFO!

- setStartVertex, no data member for EndVertex -> derived from getStartVertex of daughter particles on the fly, NULL else

Tau ID

two main approaches on the market:

TauJet: taus in hadronic events

Delphi: taus in low multiplicity events (up to ~10-15 PFOs)

plan:

release Taikan's TauJet in MarlinReco/Analysis - **DONE**

Taikan & Mikael go through details of both finders

how to combine?

wrap Delphi finder in SatoruJetFinder

lepton ID: improve by MVA, dE/dx , cluster shape

Vertexing

Vertexing is run on PandoraPFOs only

never tried on MarlinTrkTracks

need MarlinTrkTrack quality

suggestion to test:

- make basic track quality selection

- create a “TrackPFO” collection

- test vertexing on that

Taikan comits Track2PFO converter into MarlinReco/Analysis

if promissing: require Pandora to keep the relevant SOT tracks

=> work in progress by Sviatoslav / Roman /Yorgos

LCFIPlus

short-term:

- adaptive vertex finding

- soft lepton tagger using PID: put p_l in MVA

middle-term:

- Bness tagger: add “Cness”? -> after WS

- vertex mass: Graham or own π^0 reco? ->

for testing: use samples as in DBD!

enable vertex fit to read track collection directly ?

check if V0 PandoraPFO has end/start vertex correctly - has
NOT

Truth Algorithms

TrueJets

in v01-17-07

needs: fixes for Higgs in final state, ttH physsim,
gammagamma-> hadrons from Pythia

RecoMCTruthLink [to be updated]

found various missing hit-MCP relations

IMPORTANT: BeamCal hits by accident included in
PandoraPFOs in DBD production ???

fix-up will be provided

TrueShower – would it be useful? YES

Relation / Interplay with TruthVertices ???

Pair background

have file with MCParticles which go
directly through tracking volume without
backscattering ?

include pair overlay as option in stdreco?

BeamCal

not part of MarlinReco anymore, but in
new package FCalClusterer

use parametrised method

Frank will talk to Andre Sailer to understand preparation of
“TaggingEfficiency.root” input file for parametrised method

no simple fast sim parametrisation available ?

**here the path to a usable update
for ILD is still unclear**

Event-based

- FlavourTag (Taikan & Masakazu):
 - Efficiency vs rejection rate, $Z \rightarrow qq$, $ZZZ \rightarrow qqqqqq$, jet-based
 - B vs light, b vs c
 - C vs light, c vs b
 - Maybe $t\bar{t}$?
- VertexFinding (Sviatoslav & Roman):
 - Efficiency to find B / D vertex as function of
 - Number of charged particles
 - Distance from IP
 - Number of correctly assigned tracks
 - „2D colour matrix“

Event-based

- Tracking (Yorgos, Tino)
 - Efficiency and bad track rate in ttbar, mumu vs p, theta
 - With ≥ 4 Si hits ? Or ≥ 4 in VXD ? In innermost
- Particle ID in jets (Masakazu)
 - same sample as flavour tag
 - Efficiency / fake rate vs momentum, theta, ...
 - Same as single particle PID benchmarks
- Jets (Bono & Cambridge group, Lan)
 - Invariant mass of uds dijets
 - Jet energy scale
 - Residual between
 - True and reco photon energy
 - True and reco neutral hadron energy
 - True and reco charged PFO energy
 - „PFO finding efficiency / fake rate“: but based on PFOs

Single particle based

- Photons: (Daniel?, Graham)
 - Efficiency / purity vs energy, theta
 - Energy resolution, x,y,z resolution of cluster position, intrinsic cluster direction
 - Number of reco photons per true photon,
- Pi0: Graham
 - „same as photons“
 - Mass resolution
- Taus ??? : (Hieu, Taikan, Mikael)
 - „same as photons“
 - Decay mode separation
- V0, Conversions, J/Psi (Graham?)
 - Same as photon
 - Mass resolution

Single Particle based

- Particle ID: (Masakazu)
 - separately for dE/dx based, cluster-based, total
 - particles: e, mu, pi, p, K,
 - 1d histograms / matrix with probability to identify true type i as reco type j for fixed momentum: 0.5 GeV, 1 GeV, 2 GeV, ... 10 GeV
 - e/pi separation vs p etc
- Tracking (Yorgos & Tino)
 - Single mu: resolution(d0, pt) vs momentum, theta
 - Single mu efficiency vs p, theta, d0
 - Pulls for dEdx
 - FWD Tracking: included
- BeamCal
- LumiCal
- Muon system ;-)

Further Plan

patch release v01-17-07.p02 TODAY

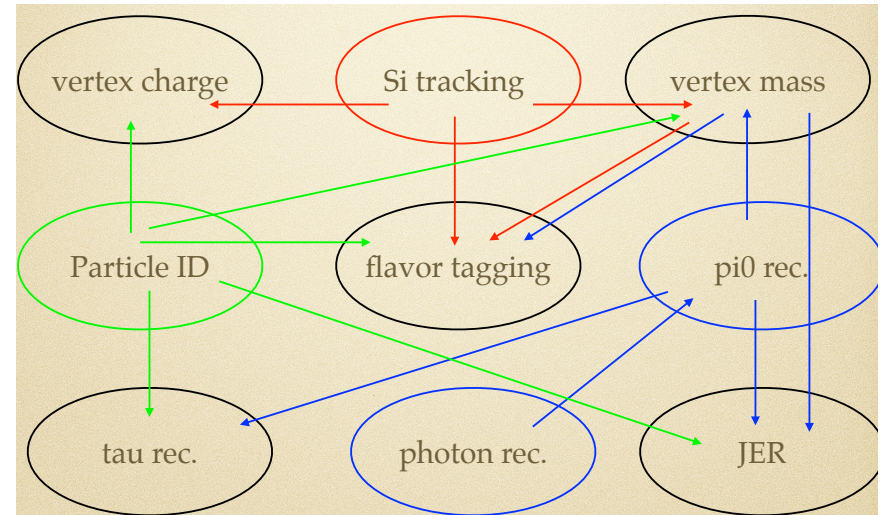
developers release v01-17-08 : next week
(before summer break)

Mokka-compatible

legacy release v01-18 (?): September ?

Conclusions – personal view

- we were **really** productive this week
- huge progress in integrated all the existing developments
- but also: significantly improved understanding of long existing stuff
- of course there remain several things to do
-> but we have a clear path to proceed!



**a big THANK YOU to all who
contributed to this intense
workshop – at DESY and
remotely!**