Summary on Simulation and Reconstruction

First ECFA Workshop on e+e- Higgs/EW/Top Factories
Hamburg 07.10.2022

Uli Einhaus

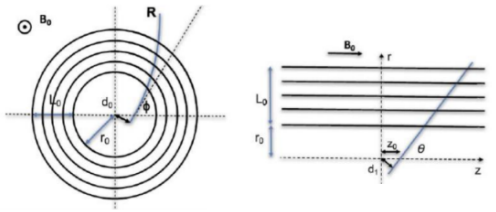
ulrich.einhaus@desy.de

Overview

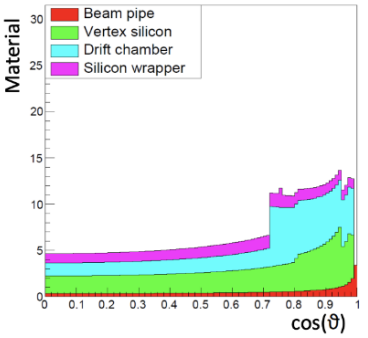
- Michele Selvaggi: Progress on Delphes
- Engin Eren: Generative Models for Calorimeter Showers
- Juraj Smieško: Reconstruction with KEY4HEP
- Nicola De Filippis: PID with Cluster Counting
- Yasser Radkhorrami: ErrorFlow - Jet Error Est. for KinFit
- Leonhard Reichenbach: Charged Lepton ID in Jets
- Uli Einhaus: A PID Framework for FHFs
- Taikan Suehara: LCFIPlus and ongoing works with DNN

Michele Selvaggi: Progress on Delphes

Link

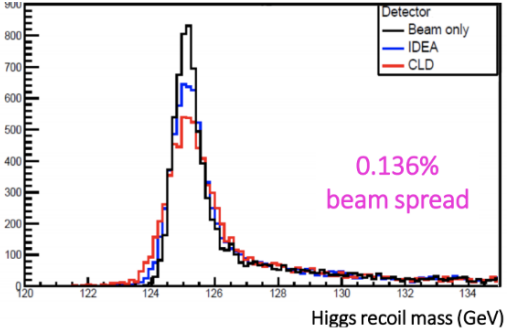


$$\vec{\alpha} = (D, \varphi_0, C, z_0, \lambda)$$



Track Smearing

- Simple tracker geometry implementation, **including material**
- Computes **full covariance matrix** (in present Delphes we have "diagonal" smearing in the 5 tracking parameters)
- Can be used for studying impact of material and realistic **HF tagging** simulation



Bedeschi, Gouskos, MS, [2202.03285]

Michele Selvaggi: Progress on Delphes

Link

```
##################
# Cluster Counting
##################

module ClusterCounting ClusterCounting {

  add InputArray TrackSmearing/tracks
  set OutputArray tracks

  set Bz $B

  ## check that these are consistent with DCHCANI/DCHNANO parameters in TrackCovariance module
  set Rmin $DCHRMIN
  set Rmax $DCHRMAX
  set Zmin $DCHZMIN
  set Zmax $DCHZMAX

  # gas mix option:
  # 0:  Helium 90% - Isobutane 10%
  # 1:  Helium 100%
  # 2:  Argon 50% - Ethane 50%
  # 3:  Argon 100%

  set GasOption 0

}
```

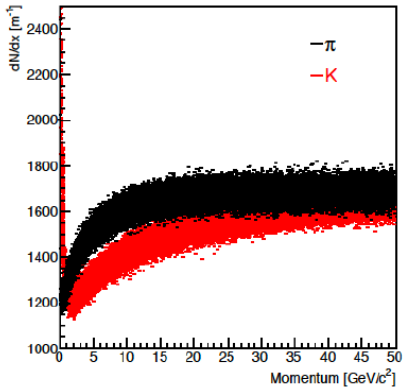
Particle Identification

- dNdx, cluster counting method, dependent on track geometry, implemented for 4 gas mixes
- Time-of-flight: implemented both for charged and neutrals

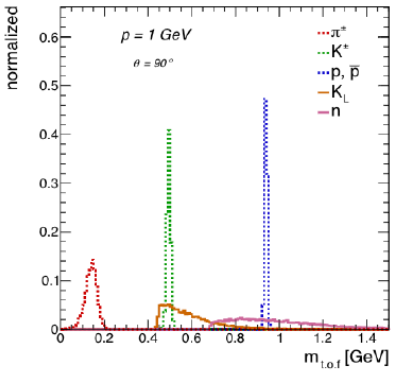
Jet Clustering

- Durham inclusive/exclusive
- Valencia

dNdx



TOF



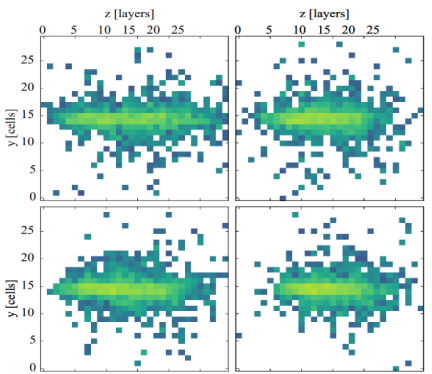
Engin Eren: Generative Models for Calorimeter Showers

Link

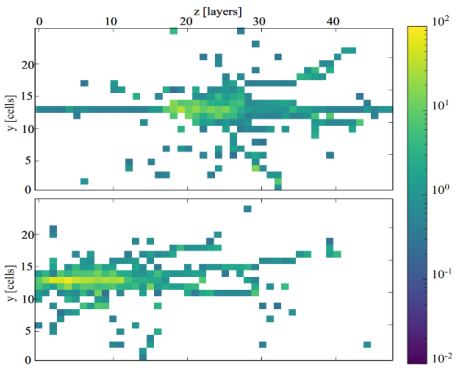
- MC simulation is computationally intensive
  - Calorimeters most intensive part of detector simulation
- **Generative models** potentially offer orders of magnitude speed up
- For the first time more challenging hadron showers in a high granular hadronic calorimeter

Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	2684 ± 125	$\times 1$
	WGAN	47.923 ± 0.089	$\times 56$
	BIB-AE	350.824 ± 0.574	$\times 8$
GPU	WGAN	0.264 ± 0.002	$\times 10167$
	BIB-AE	2.051 ± 0.005	$\times 1309$

*WGAN: (Wasserstein-) Generative Adversarial Neural Network
*BIB-AE: Bounded-Information Bottleneck Autoencoder

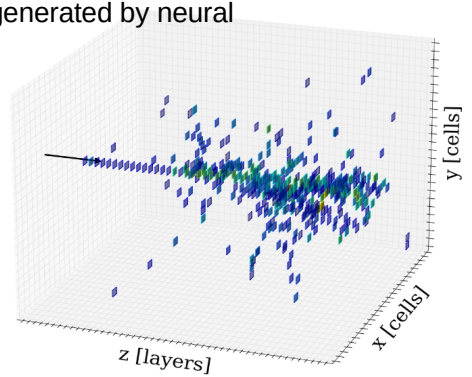


Cell Layout: 30 x 30 x 30 (Photon showers)



Cell Layout: 48 x 25 x 25 (Charged pion showers)

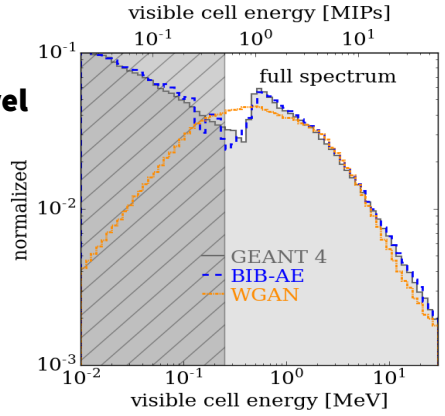
Shower generated by neural network!!

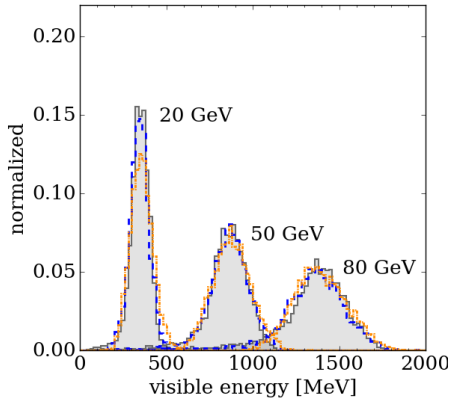


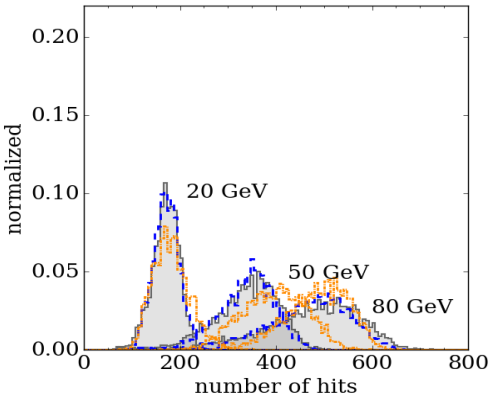
Engin Eren: Generative Models for Calorimeter Showers

Link

Generator Level

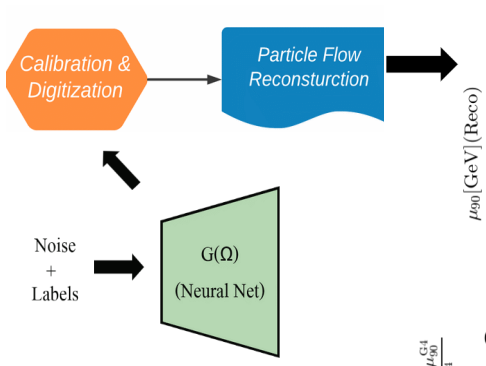


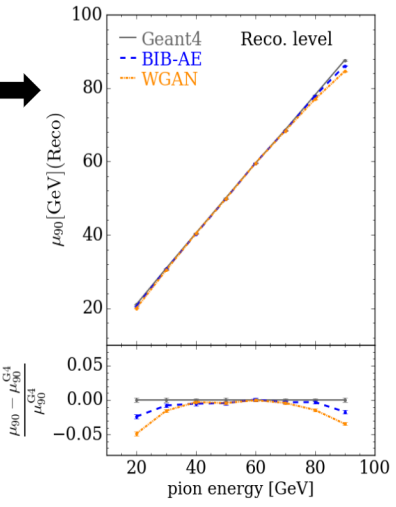


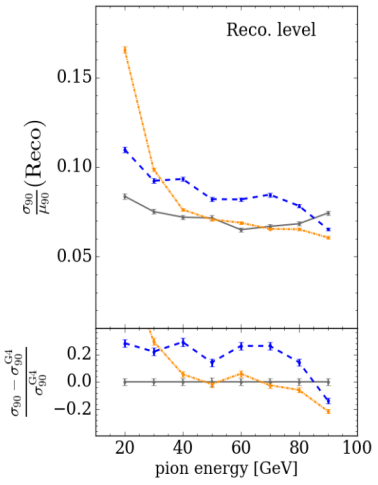


arXiv:2112.09709

Reco Level





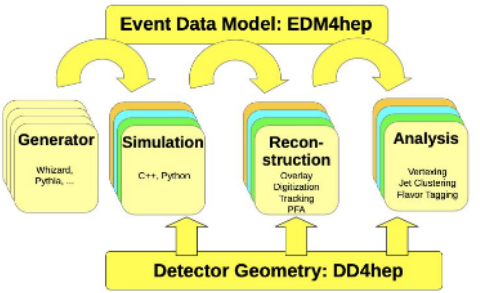


- Great progress
- Publications ongoing
- Various open questions

Juraj Smieško: Reconstruction with KEY4HEP

Link

- Key4hep: Gaudi, EDM4hep, DD4hep, Spack
- Key4hep has ability to integrate other advanced reconstruction tools/frameweworks
- `k4MarlinWrapper` helps to bridge transitional period
  - DDMarlinPandora, LCFIPlus, ConformalTracking, …
- Integration of large frameworks underway
  - K4CLUE, k4Pandora, k4ActsTracking
- Effort required to port reconstruction of already existing detector concepts to Key4hep

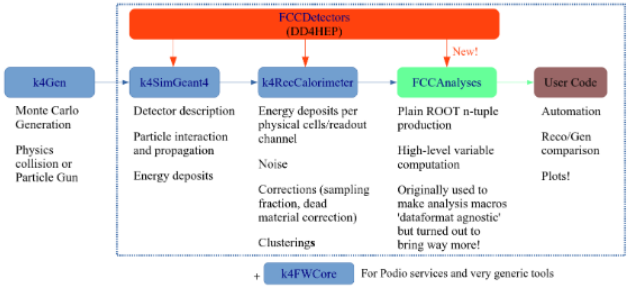


Juraj Smieško: Reconstruction with KEY4HEP

Link

LAR CALORIMETER

- Sampling Calorimeter based on LAr/LKr + Pb/W
- Simulation/Reconstruction fully steered in Gaudi
- Several Gaudi based algorithms include
  - Sampling fraction determination
  - Upstream/Downstream energy correction
  - Adding noise to Calo Cells
  - Clustering: Sliding Window or TopoCluster based

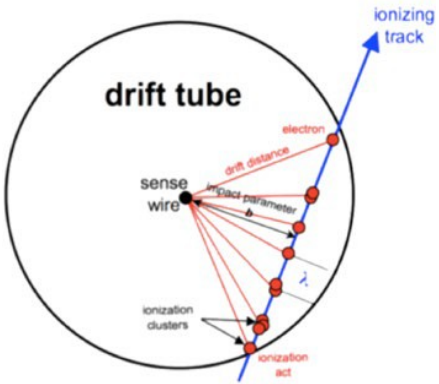


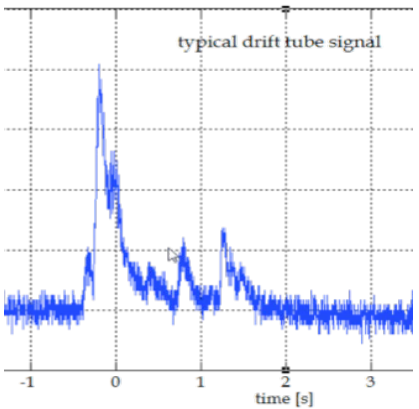
CLD

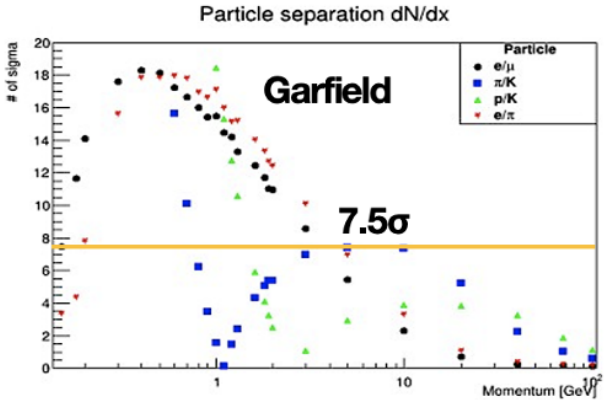
- Uses DDSim to simulate events
- Heavy use of the converters
- The reconstruction consists of
  - Background Overlay, Digitization
  - Track Pattern Reconstruction (ConformalTracking), track fit
  - Particle Flow Reconstruction (PandoraPFA)
  - Vertexing and Flavour Tagging (LCFIplus)
  - Full CLD reconstruction in gaudi
- Input and output are in EDM4hep

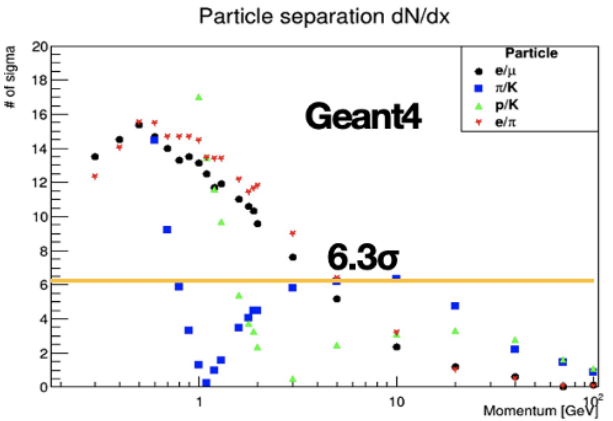
Nicola De Filippis: PID with Cluster Counting

Link









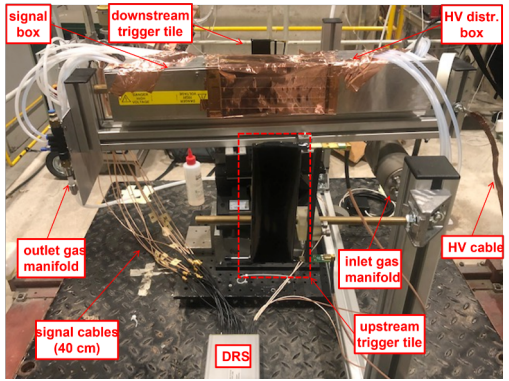
dE/dx: Truncated mean cut (70-80%) reduces the amount of collected information. n = 112 and a 2m track at 1 atm give $\sigma \approx 4.3\%$

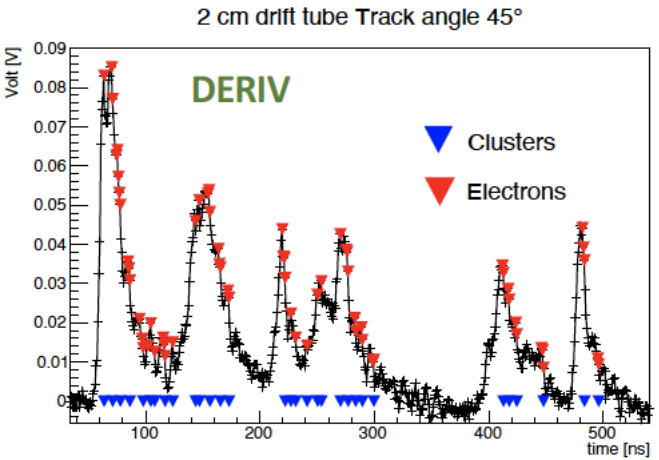
dN_{cl}/dx: δ_d = 12.5/cm for He/iC_4H_{10}=90/10 and a 2m track gives $\sigma_{dNcl/dx}/(dN_{cl}/dx) = N_{cl}^{-1/2} < 2.0\%$

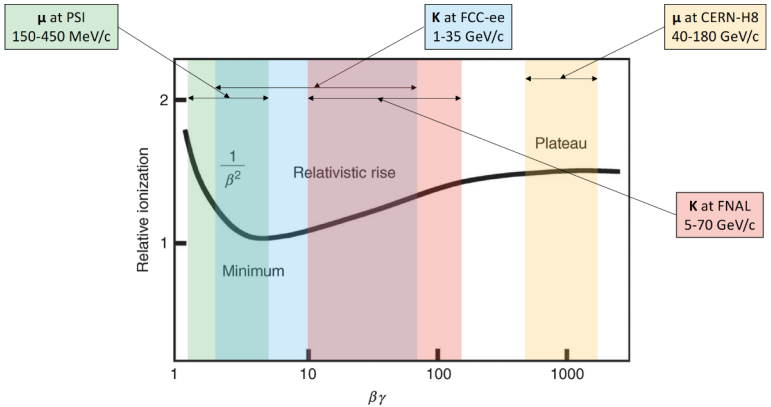
- Garfield++ in agreement with analytical calculations up to 20 GeV/c, then falls much more rapidly at higher momenta
- The particle separation in GEANT4 is considerably worse than in Garfield++. Why?
- Need to check with data → beam test!

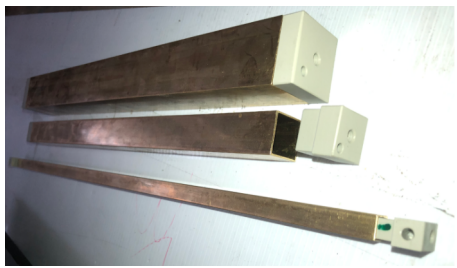
Nicola De Filippis: PID with Cluster Counting

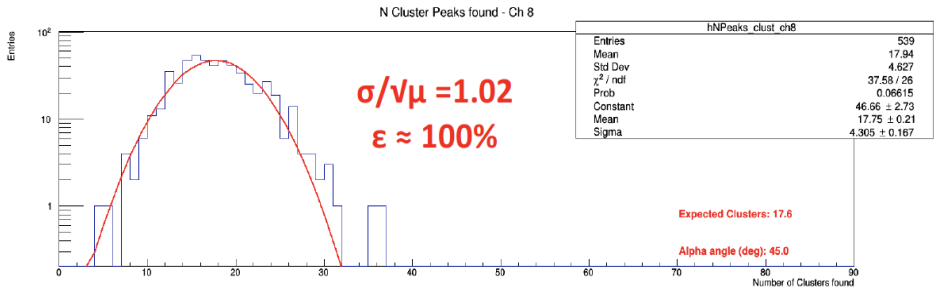
Link











- Beam test at CERN-H8 with 165 GeV μ
- Poissonian behaviour
- Meaurements and predictions about the number of clusters are in very good agreement, with 1cm cell size

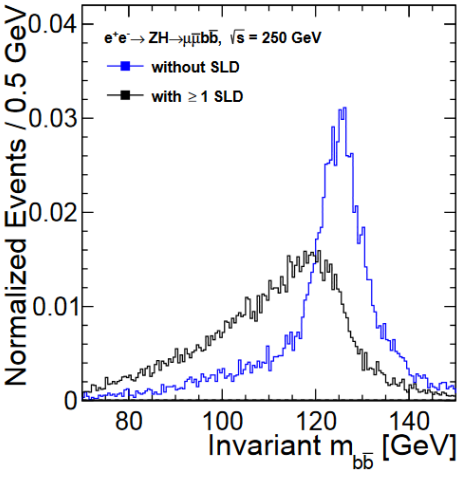
- Study the 1/ region at PSI (βγ ≈ 1–4)
- Study the relativistic rise at MT6/Fermilab (βγ ≈ 10–140)

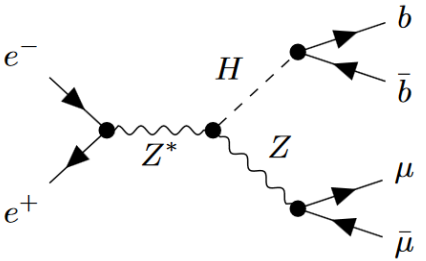
Yasser Radkhorrami: ErrorFlow - Jet Error Est. for KinFit

Link

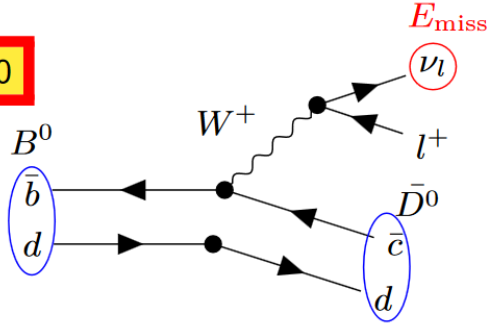
- Number of B-/C-hadron semil-eptonic decays (SLD) in $e^+e^- \to ZH \to \mu\bar{\mu}b\bar{b}$ events

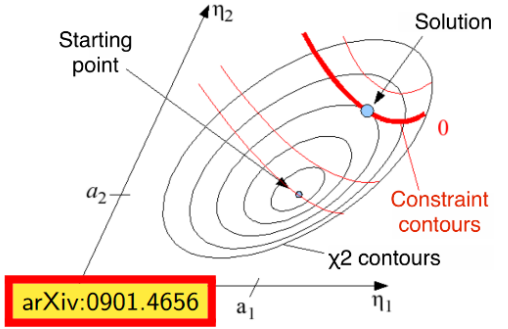
		nBSLD		
		0	1	2
nCSLD	0	34%	24%	4%
	1	18%	12%	2%
	2	3%	2%	0%





arXiv:2105.08480





arXiv:0901.4656

- retrieve missing momentum (up to sign ambiguity) from event and decay kinematics in a highly granular detector.
- **Resolve sign ambiguity using overall event kinematics ⇒ kinematic fit!**

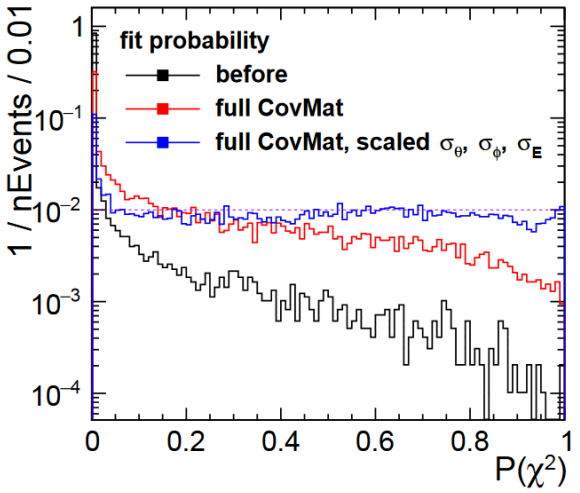
Exploit well-known initial state in e^+e^- colliders

⇒ **need error parametrization, in particular for jets**

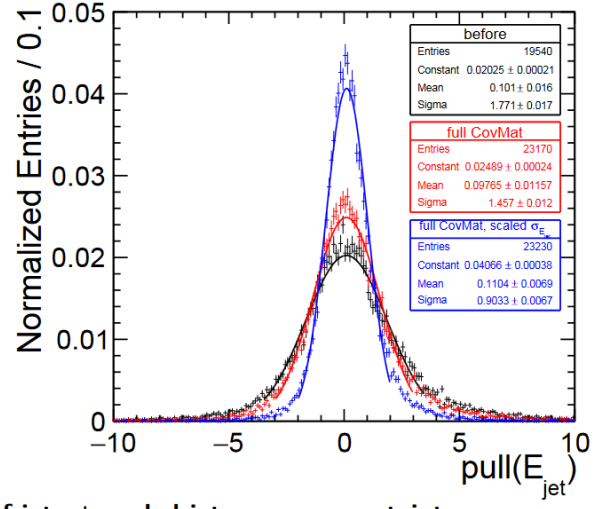
Yasser Radkhorrami: ErrorFlow - Jet Error Est. for KinFit

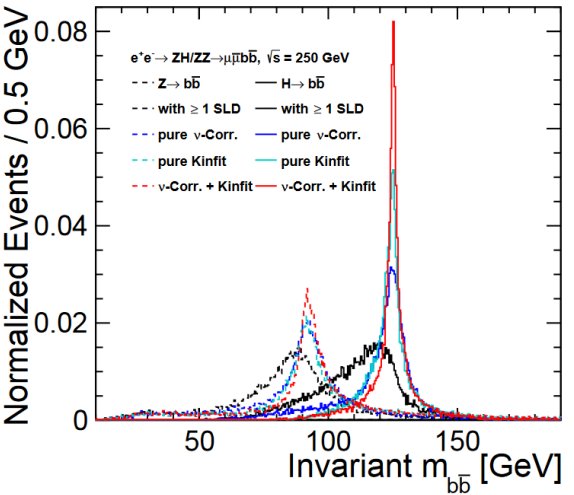
Link

- fit probability



- pull distribution





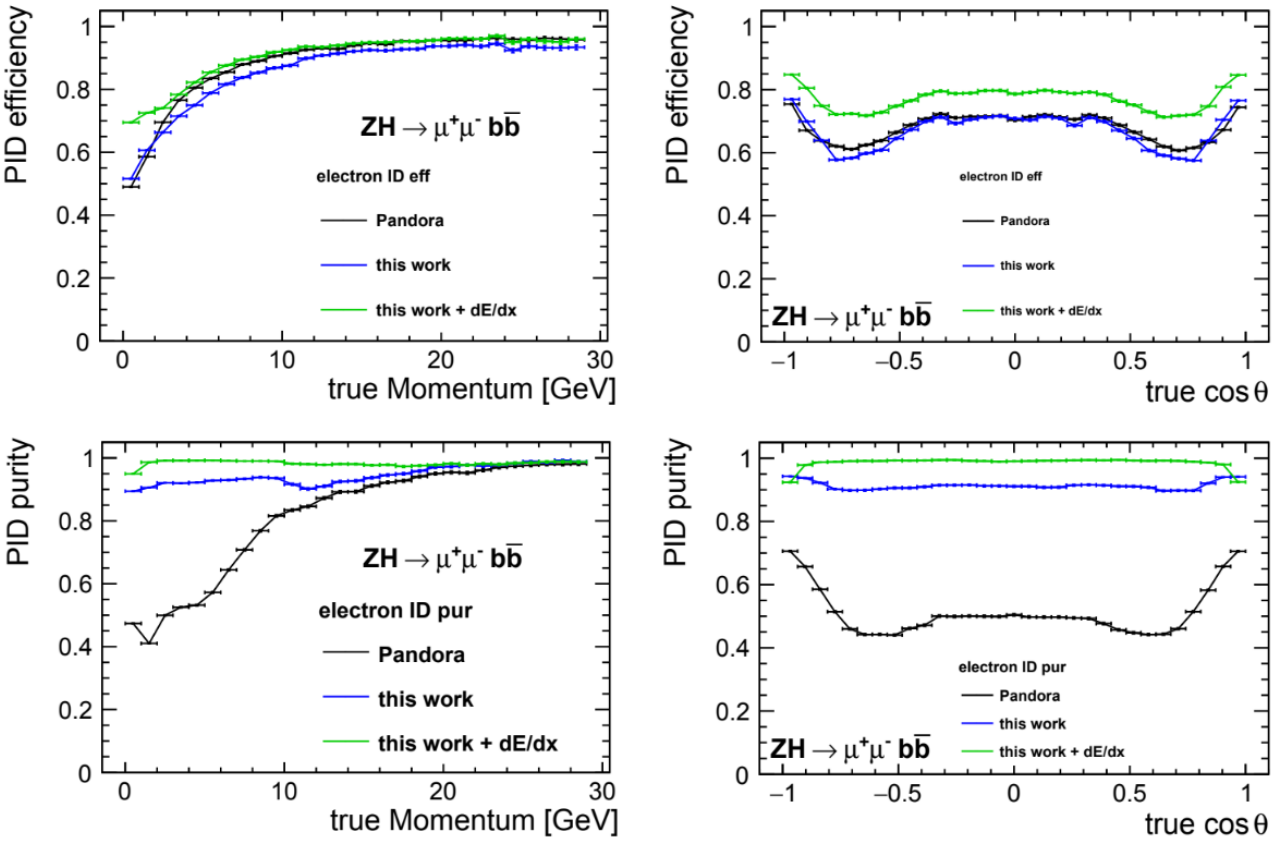
Improved kinematic fit performance with full CovMat of jets + scaled jet energy uncertainty

Leonhard Reichenbach: Charged Lepton ID in Jets

Link

- Electron and muon ID so far studied for isolated particles, e.g. Z→e+e-/μ+μ-
- First look into e/μ in H→bb for neutrino correction
  - Black: raw particle flow
  - Blue: incl. detailed cluster shape
  - Green: blue + dE/dx

dE/dx gives important contribution to e-ID up to 20 GeV

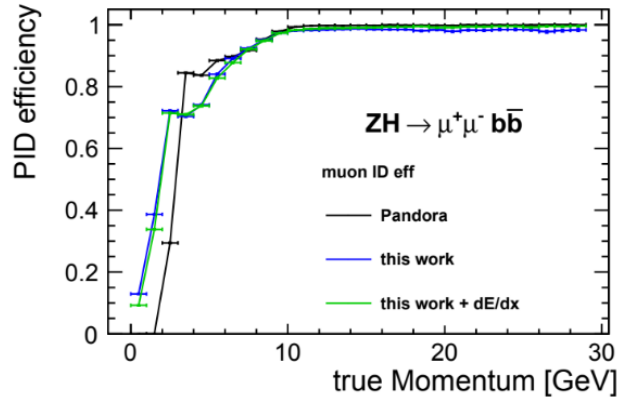


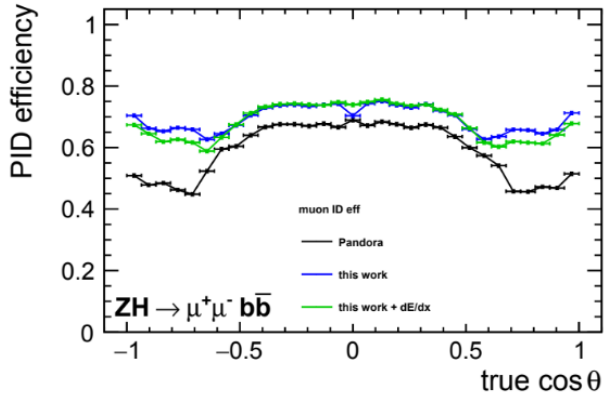
Leonhard Reichenbach: Charged Lepton ID in Jets

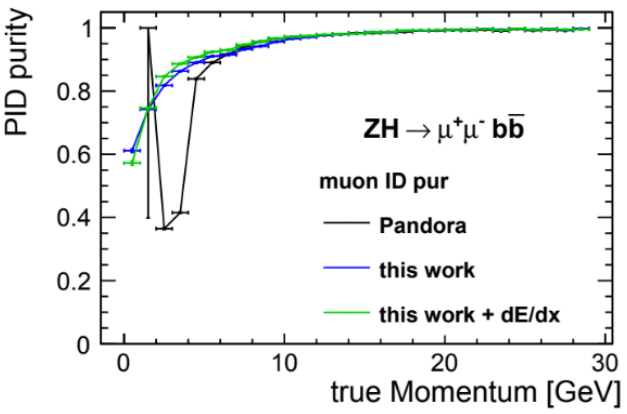
Link

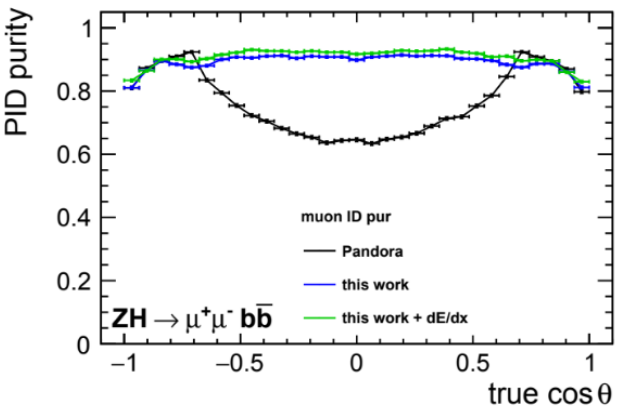
- Electron and muon ID so far studied for isolated particles, e.g. Z→e+e-/μ+μ-
- First look into e/μ in H→bb for neutrino correction
  - Black: raw particle flow
  - Blue: incl. detailed cluster shape
  - Green: blue + dE/dx

dE/dx gives important contribution to e-ID up to 20 GeV









Uli Einhaus: A PID Framework for FHFs

Link

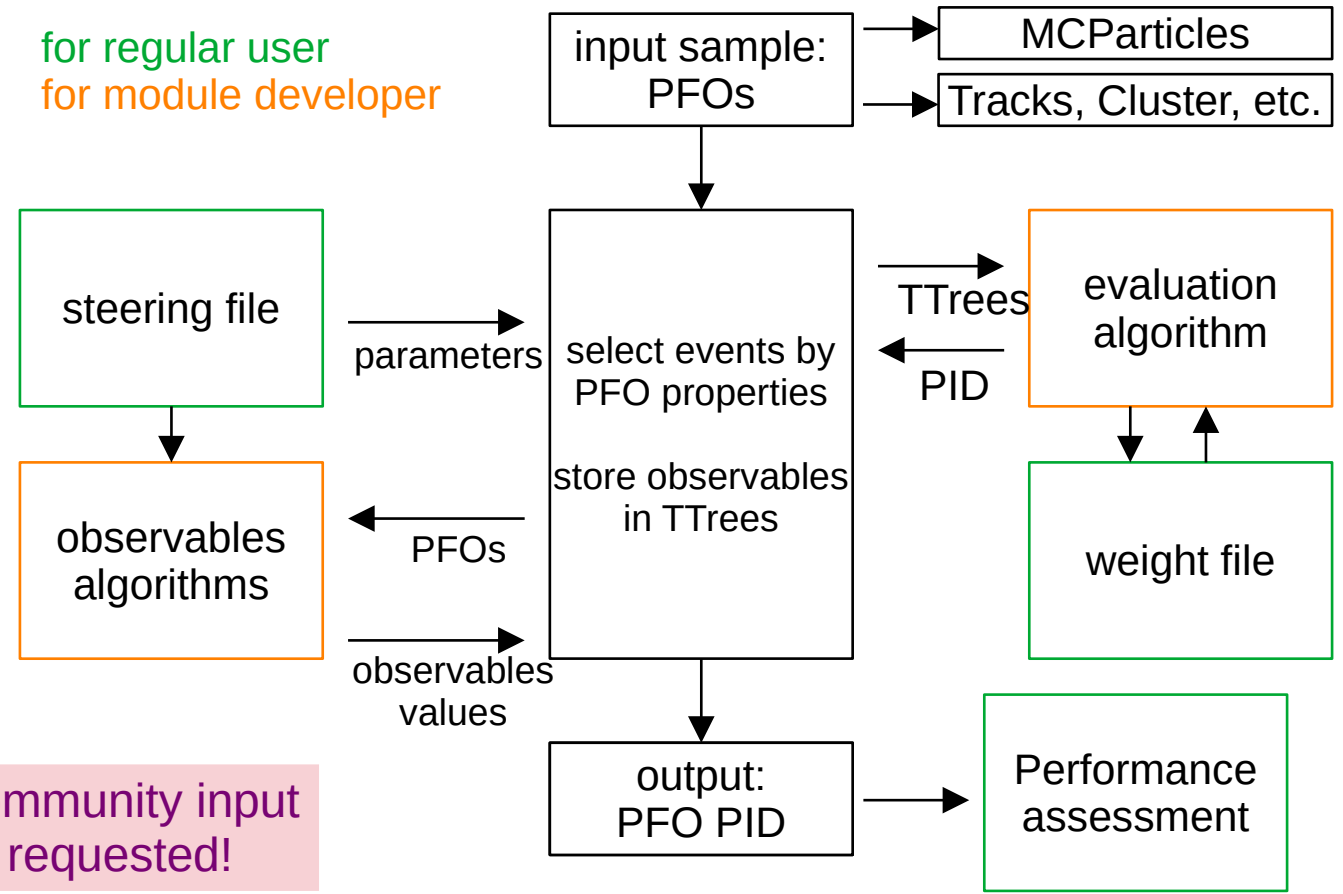
- Optimise and compare detectors
  - Is Silicon dE/dx worth considering?
  - At what timing resolution starts TOF to be relevant for flavour tagging?
  - How does **my** physics result depend on the dE/dx resolution?
  - What if we add a RICH to SiD?

- Proposal: ‚Comprehensive PID Processor'
- For now, being implemented in LCIO / Marlin
  - immediately usable in Key4HEP via 'Marlin wrapper'
  - target: implement in EDM4HEP, make available to whole FHFs community
- Modular:
  - observables algorithms
  - training methods / evaluation algorithms
- Core code takes care of book keeping
  - simple, well defined data structures for storage (TTree) and interfaces (std::vector)

Uli Einhaus: A PID Framework for FHFs

Link

- Steering file
  - input sample
  - observables algorithms
  - signal categories PDGs
  - evaluation algorithm
  - weight file
  - sample cuts etc.
- Possibility to store observables and do training separately (python etc.)
- Separate performance assessment for various plots incl. eff./pur. MC PDG vs. Reco PDG and separation power



Taikan Suehara: LCFIPlus and ongoing works with DNN

Link

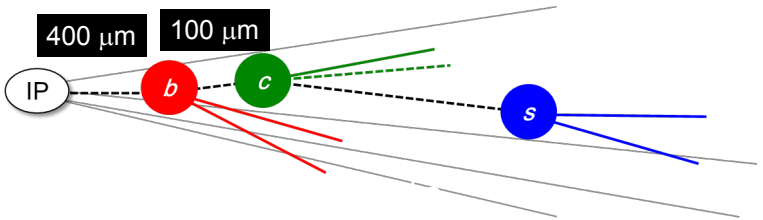
https://doi.org/10.1016/j.nima.2015.11.054
https://github.com/lcfiplus/LCFIPlus

Standard flavor tagging software for ILD/CLIC since 2013

Integrated software for

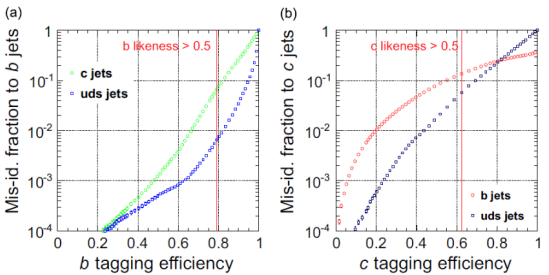
- Vertex finding (tear-down, build-up)
- Jet clustering (incl. beam rejection)
- Flavor tagging (b/c/q separation) based on BDT/TMVA

Data/process model independent of LCIO/Marlin structure

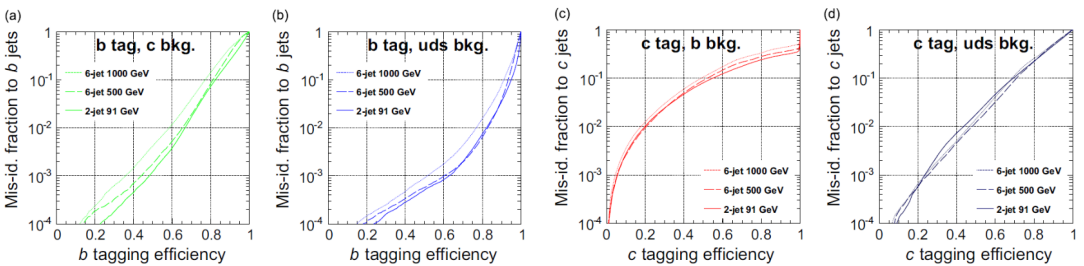


Performance of vertex finding

(#vtx, #pseudo-vtx)	b jet (%)	c jet (%)	uds jet (%)
(0, 0)	21.3	59.3	98.1
(0, 1)	1.61	0.17	0.01
(1, 0)	39.7	39.8	1.80
(1, 1)	13.5	0.54	0.02
(2, 0)	23.8	0.19	0.04



Performance of b/c tagging with Z → qq 91 GeV and 6q 1000/500 GeV



Taikan Suehara: LCFIPlus and ongoing works with DNN

Link

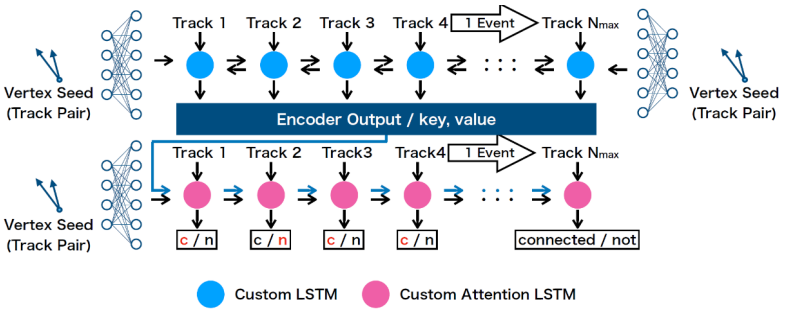
Vertex finding / Flavor tagging

Vertex finder with LSTM

- Following LCFIPlus strategy
  - Replacing pair selection and track association with NN and LSTM
- Similar performance to LCFIPlus

Flavor tagging with GNN ongoing

* (Deep) Neural Network with Long Short-Term Memory



Network architecture for track association

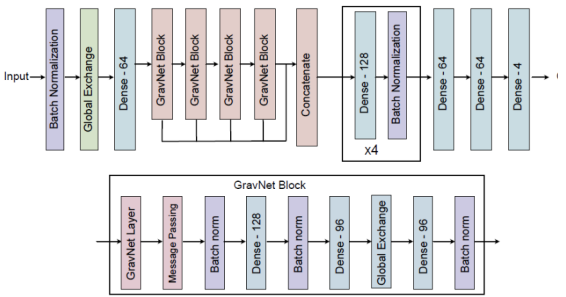
Calo clustering / PFA

Introducing GravNet / object condensation (from HGCal) to ILC

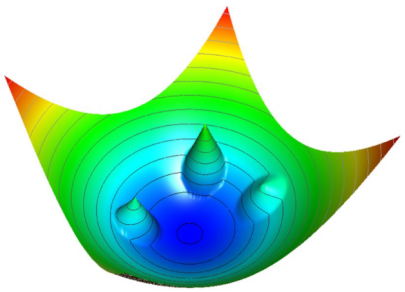
GravNet: contrastive GNN layers

Object condensation: to gather true hits

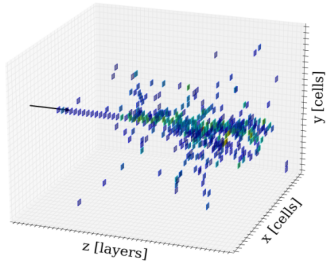
First implementation done: performance evaluation ongoing

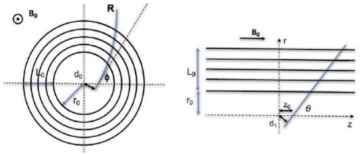


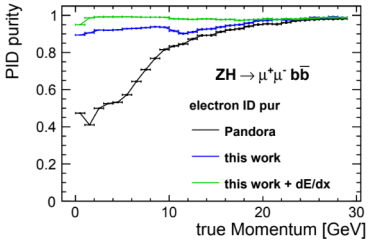
GravNet structure



Object condensation

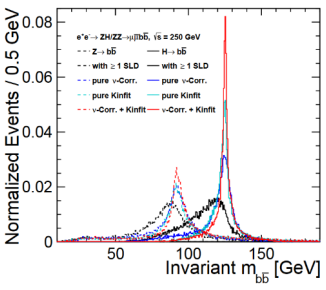


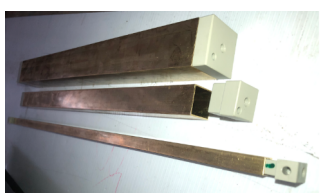


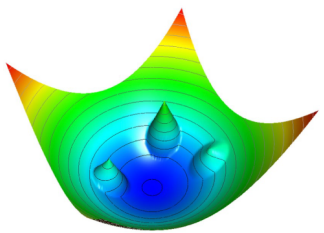


Lots of work done, lots of work ahead of us!

Thanks to all contributors!







Community input requested!