Progress on Reconstruction

An overview of recent developments

Uli Einhaus
Second ECFA Workshop on e^+e^- HTE Factories
12.10.2023 Paestum
ulrich.einhaus@desy.de

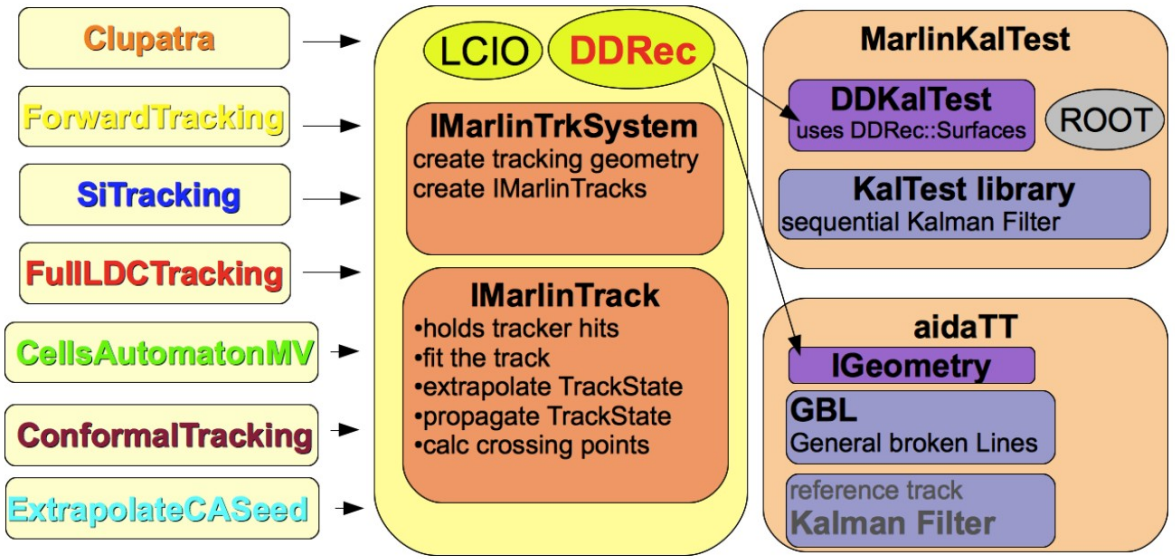
Disclaimer & Overview

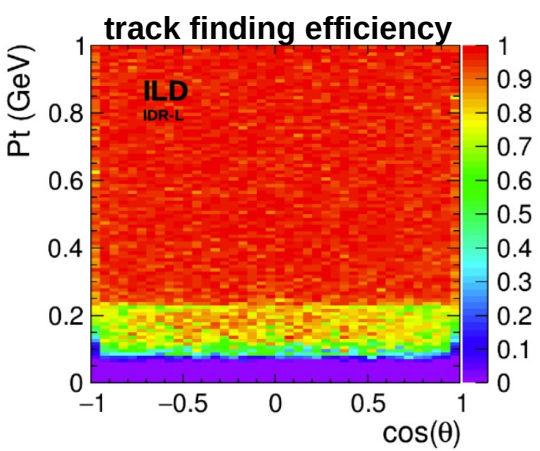
- Biased selection
- Based on First and Second ECFA Workshops on Reconstruction
- Slides stolen from various contributions from many people there and elsewhere

- Highlight recent developments in
  - Tracking
  - Calorimetry / Particle Flow
  - Particle Identification
  - Flavour Tagging
  - High Level Reconstruction

MarlinTrk

- Generic, modular tracking toolkit
- Separate track finding for TPC, Silicon and forward systems
- Combined track fitting using Kalman filter
- Conformal tracking implemented
- Looking to interface with ACTS

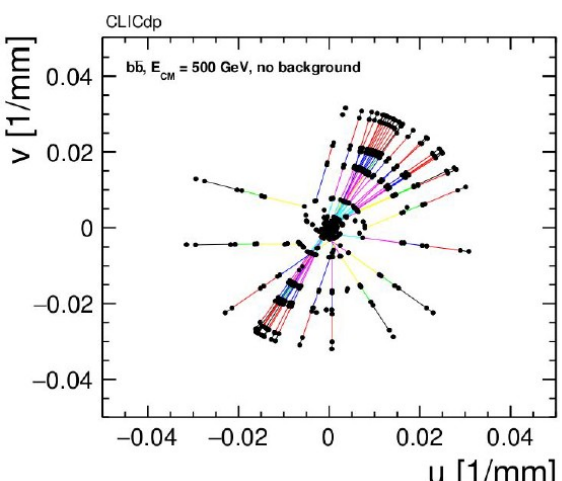




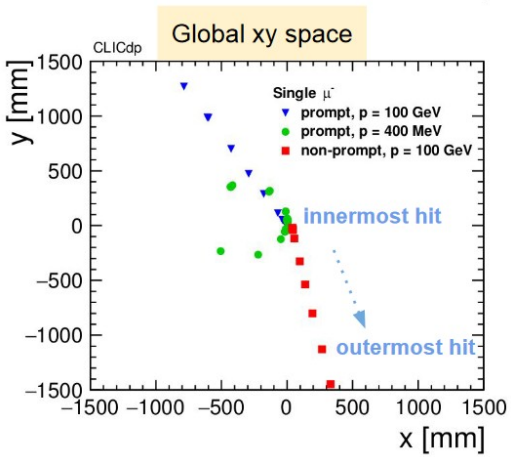
https://indico.cern.ch/event/1124095/#3-marlintrk-and-tracking-in-il

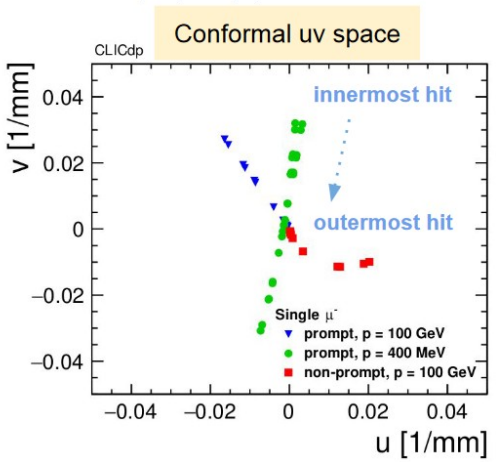
Conformal Tracking

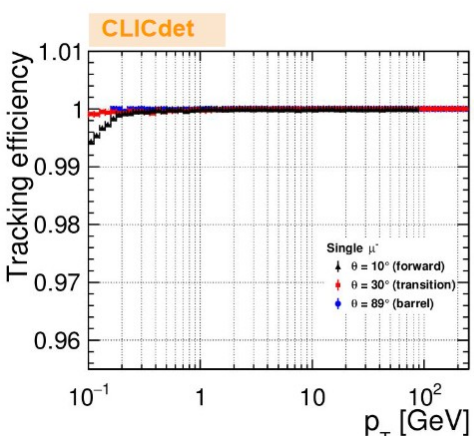
- Developed for all-Silicon trackers
- ‘Inverts’ geometrical space to find tracks via cellular reconstruction and linear fits to circular tracks
- Robust and flexible
- Integration in Key4hep and validation package make usage simple

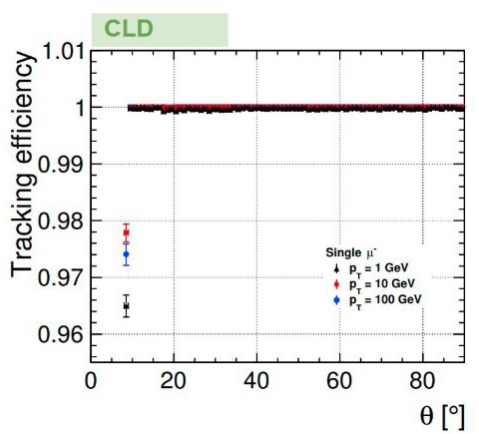


$u = x / (x^2 + y^2)$ $v = y / (x^2 + y^2)$





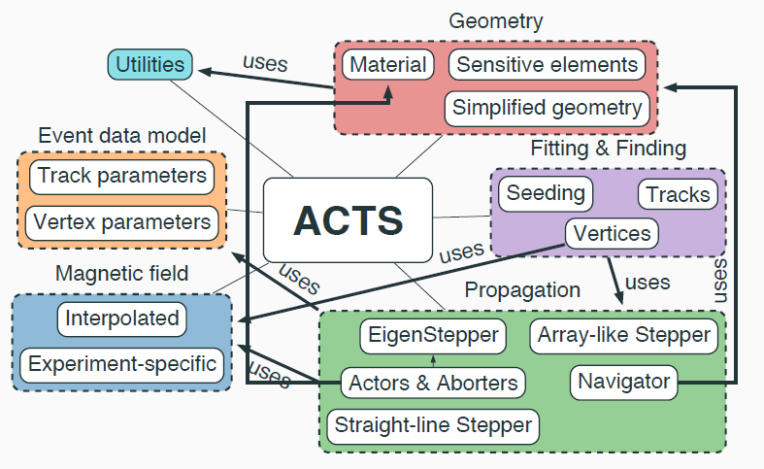


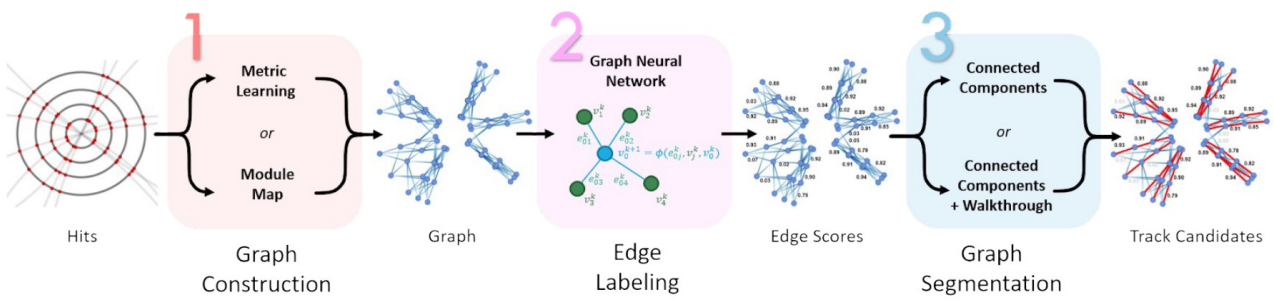


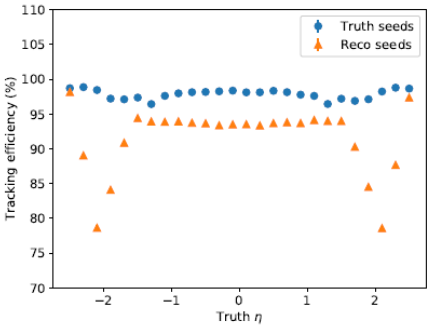
https://indico.cern.ch/event/1124095/#4-conformal-tracking-sid-clic

ACTS

- A Common Tracking Software, independent platform for tracking
- Kalman-based fitter, thread-safe, efficient
- Novel aspects:
  - Gaussian sum filter approximates non-Gaussian noise via Gaussian mixture
  - Work ongoing on Graph Neural Network for track finding
- Looking to integrate in Key4hep





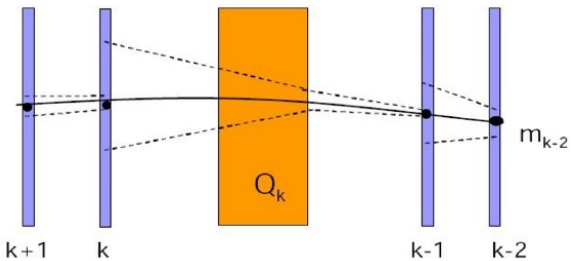


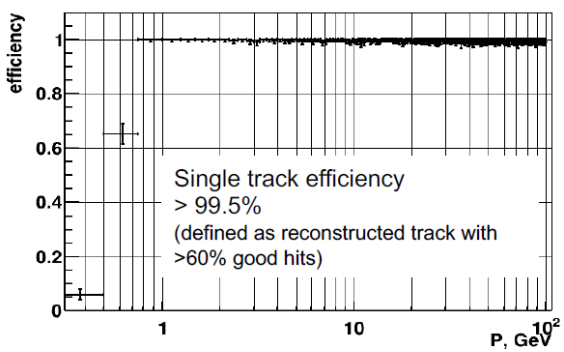
↑ $t\bar{t}$ events, $\langle\mu\rangle = 200$, TrackML geo., 2 T solenoid field

https://indico.cern.ch/event/1124095/#6-acts

Drift Chamber Tracking

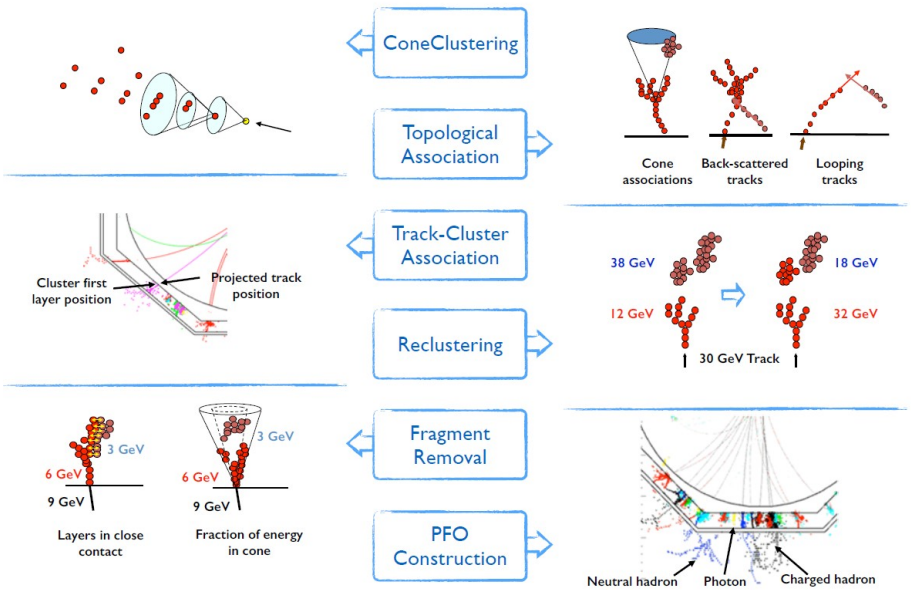
- IDEA drift chamber stereo geometry with 1-dim. wires different from 0-dim. space points in Si or TPC
- Track finding / pattern recognition:
  - local approach, starting with seed, propagate through detector
  - global approach, treating hits simultaneously
- For Kalman fitter looking into genFit2 and ACTS
- Work ongoing for implementation in Key4hep

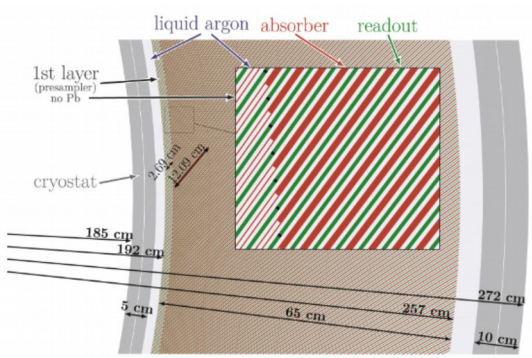


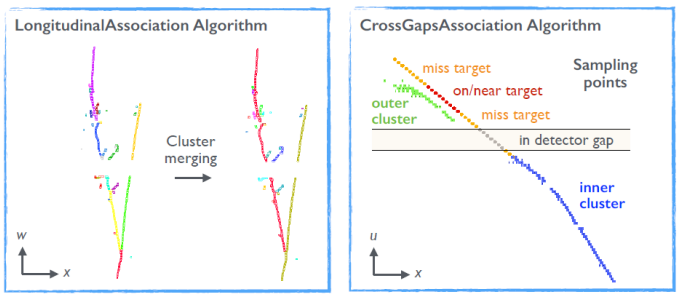


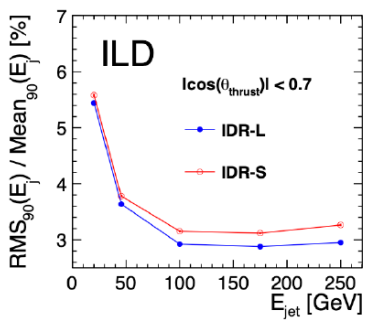
Pandora

- Established particle flow framework, using a number of algorithms for different steps of clustering and topological analysis
- Novel:
  - application to LAr TPC
  - application to ALLEGRO LAr Calo
  - implementation in Gaudi







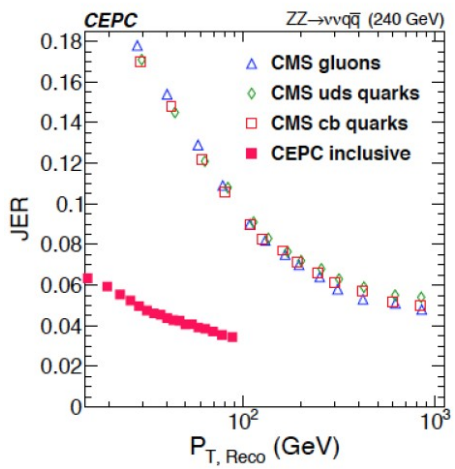


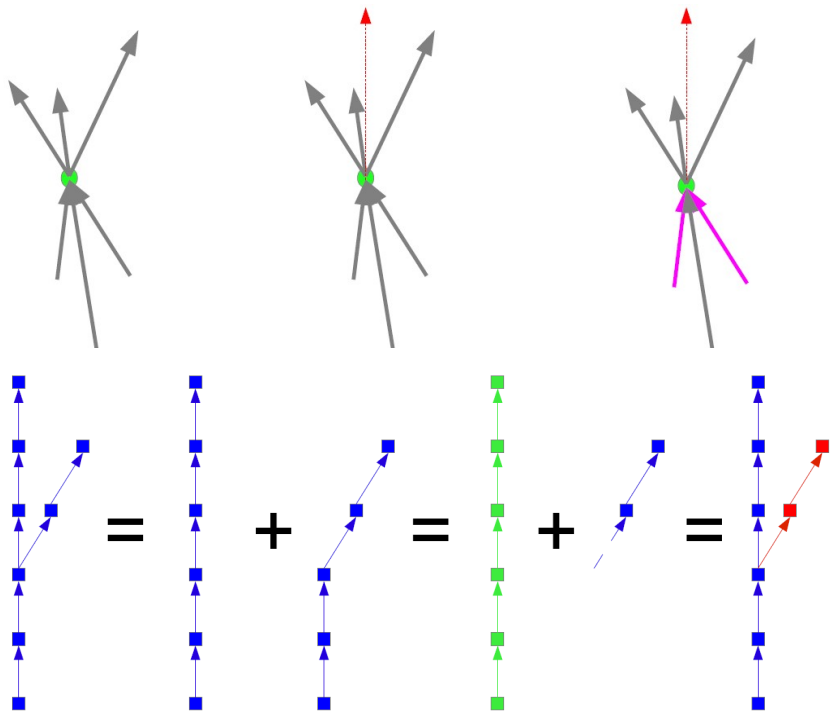
https://agenda.infn.it/event/34841/contributions/207745/

https://indico.cern.ch/event/1124095/#9-particle-flow-pandora

Arbor

- ‘Tree-building’ via hits to connectors and connectors to branches
- Extensively used for CEPC Baseline (ILD) and IDEA, fulfilling physics requirements

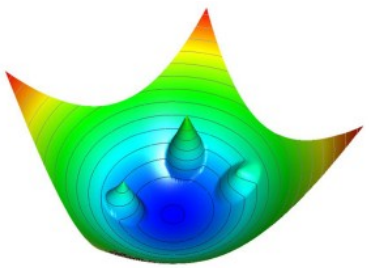


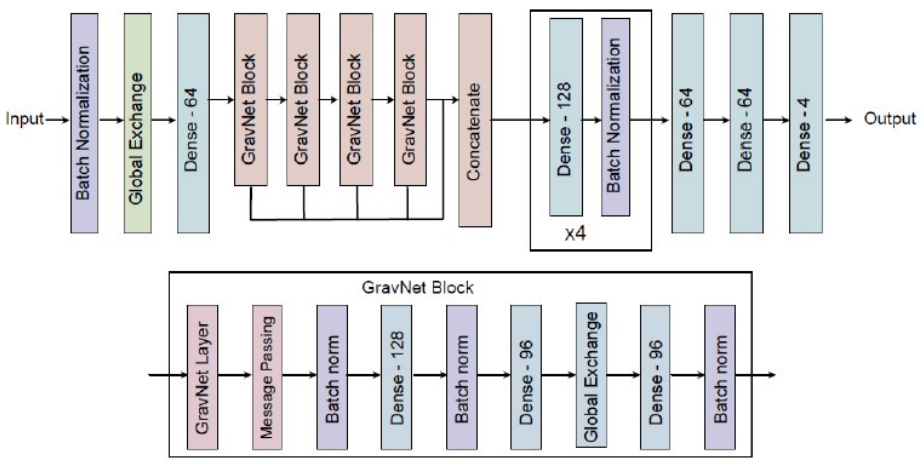


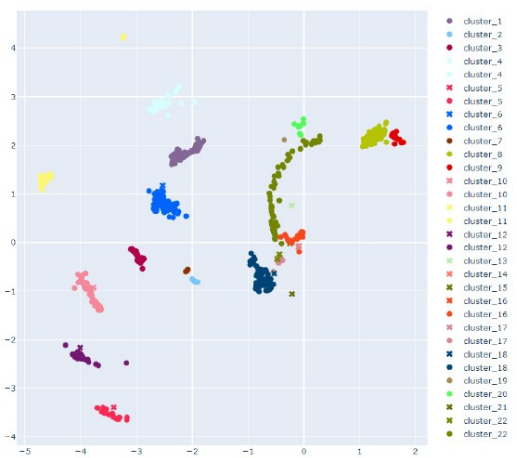
https://indico.cern.ch/event/1124095/#10-particle-flow-arbor

GNN Particle Flow

- Uses DNN-based GravNet with ~30 hidden layers, with object condensation, i.e. largest hits work as condensation points of clusters
- Promising initial results







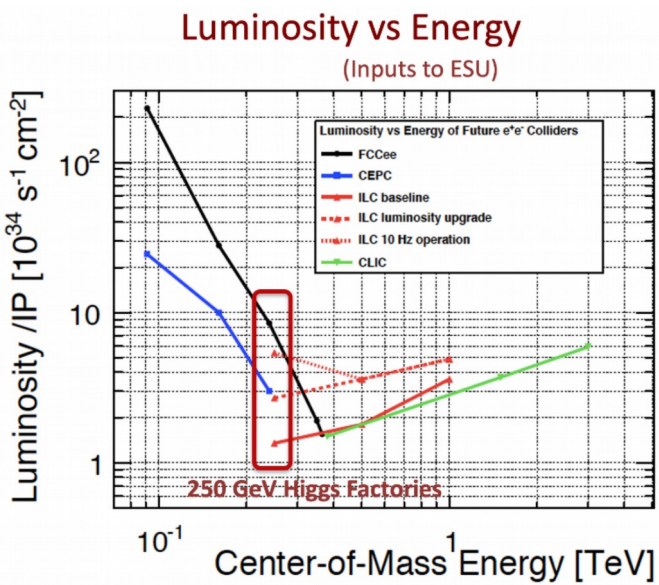
https://agenda.infn.it/event/34841/contributions/208239/

PID & Flavour Tagging

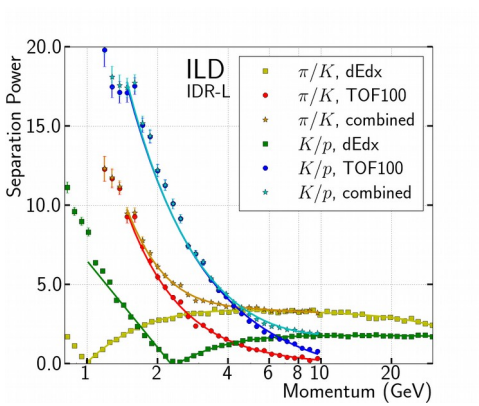
- Advent of circular collider proposals, focus of European Strategy on FCC

  → Z pole, W threshold: PID more effective, flavour physics more interesting

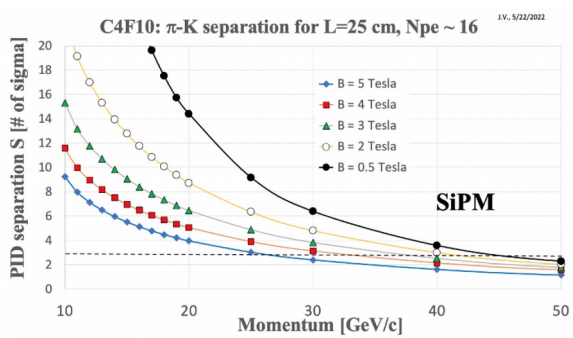
  → increased interest in these studies



https://arxiv.org/abs/1903.01629



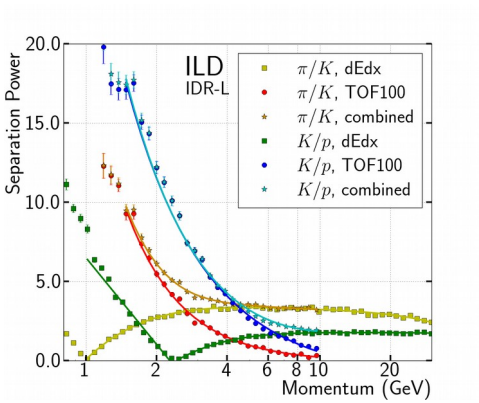
https://arxiv.org/abs/2003.01116



https://arxiv.org/abs/2203.07535

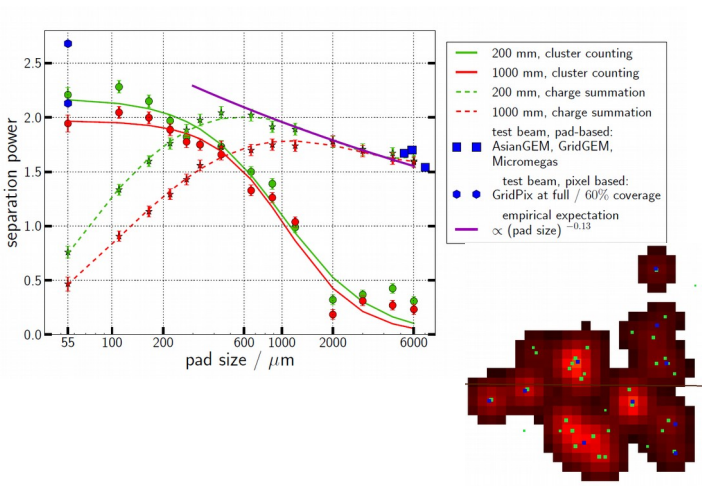
Specific Energy Loss

- ILD Time Projection Chamber
- Measurement of dE/dx: reconstruct geant4-based full-simulation ionisation
- Dedicated simulation shows potential of high granularity PixelTPC for enhanced dE/dx (30-40% higher performance) and possibly cluster counting (dN/dx)
- CEPC (ILD-based) TPC coming to similar results



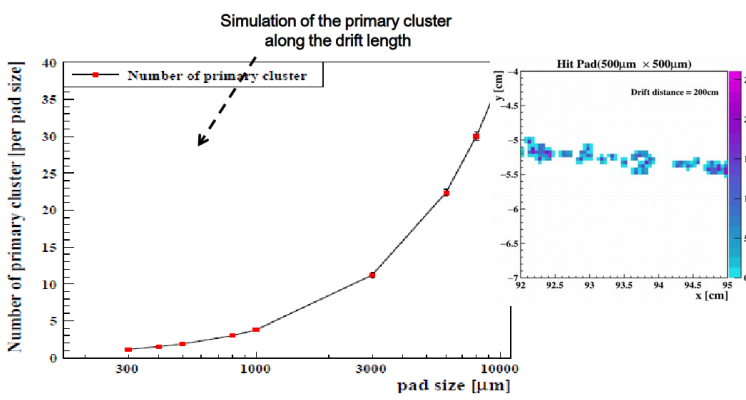
https://arxiv.org/abs/2003.01116

LCTPC



https://arxiv.org/abs/2205.12160

CEPC

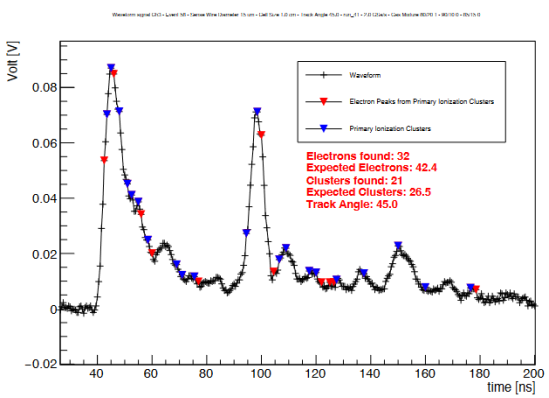


https://indico.desy.de/event/34916/contributions/147134/

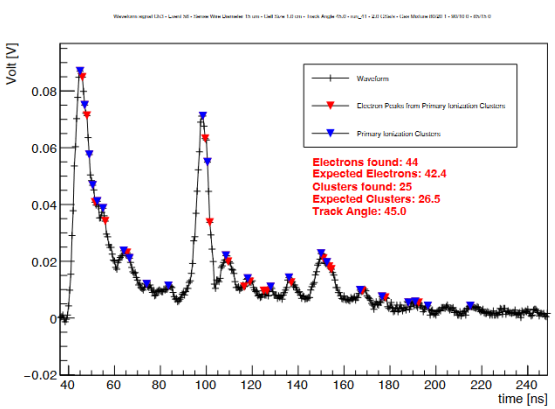
Specific Energy Loss

- IDEA Drift Chamber, Bari group
- Simulation in geant4 and Garfield, compared against test beam
- Measurement of dN/dx, i.e. cluster counting dN/dx (~ factor 2 better than dE/dx) via timing, testing 2 algorithms to extract number of clusters from the signal
  - Derivative algorithm: scan through signal in small steps and use 1st and 2nd derivative to determine peak
  - Running template algorithm: template fit of experimental pulse shape, cut on χ^2

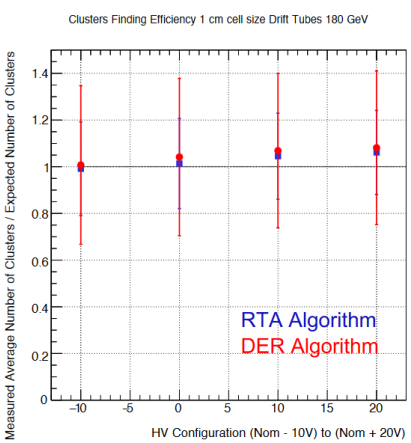
DERIV Algorithm



RTA Algorithm



1 cm drift tubes



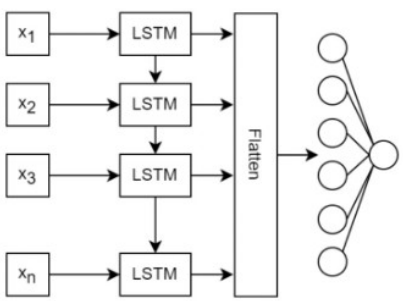
https://agenda.infn.it/event/34841/contributions/208865/

Specific Energy Loss

- IDEA Drift Chamber, IHEP group
- Approach with RNN-LSTM for peak finding and DGCNN for clustering

Peak finding with LSTM

Why LSTM? Waveforms are time series

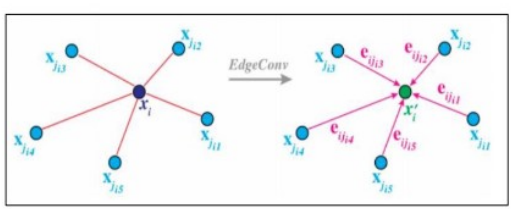


- Architecture: LSTM (RNN-based)
- Method: Binary classification of signals and noises on slide windows of peak candidates

LSTM: Long Short-Term Memory

Clusterization with DGCNN

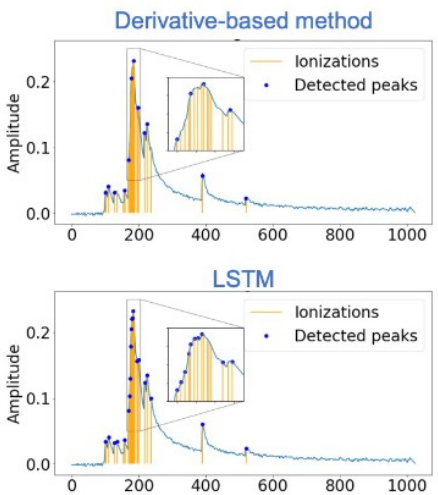
Why DGCNN? Locality of the electrons in the same primary cluster, perform massage passing through neighbour nodes in GNN

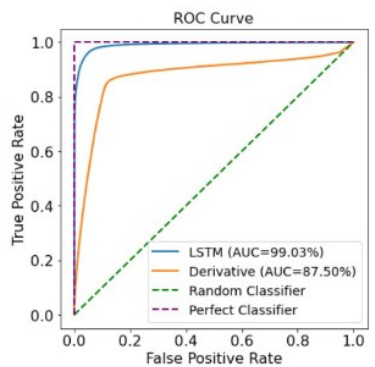


arXiv: 1801.07829

- Architecture: DGCNN (GNN-based)
- Method: Binary classification of primary and secondary electrons

DGCNN: Dynamic Graph Convolutional neural networks





LSTM model is better classifier compared to derivative-based model

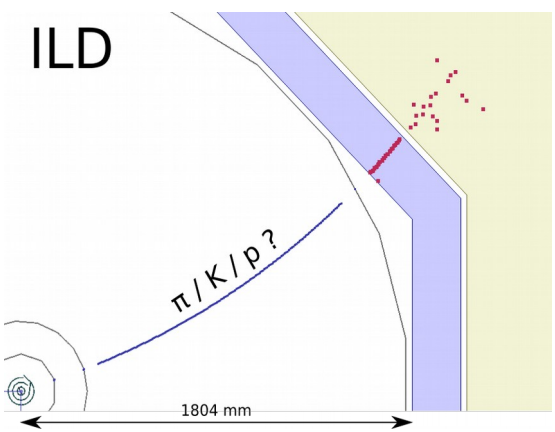
https://agenda.infn.it/event/34841/contributions/208865/
https://indico.ihep.ac.cn/event/20669/#3-peak-finding-algorithm-for-c

Time of Flight

- 30ish-ps timing in Silicon for LHC pile-up rejection can be used for low-mom PID
- Mathematically simple to implement a first estimate with a given timing T precision
  - included in DELPHES
- In ILD sim/reco based on calorimeter hits, different algorithms

$$m = p\sqrt{\frac{c^2T^2}{L^2} - 1}$$

  - 'full' reconstruction implemented with reconstructed harmonic means of track length L and momentum p

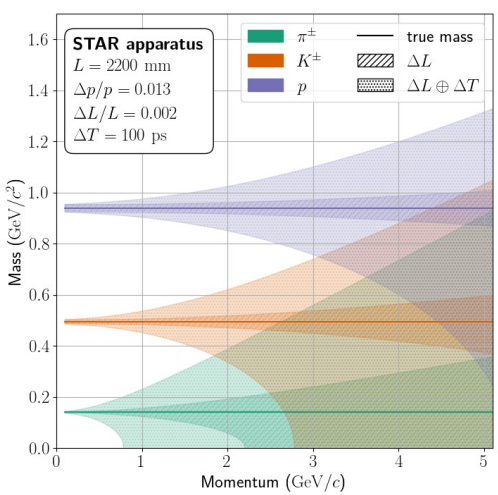


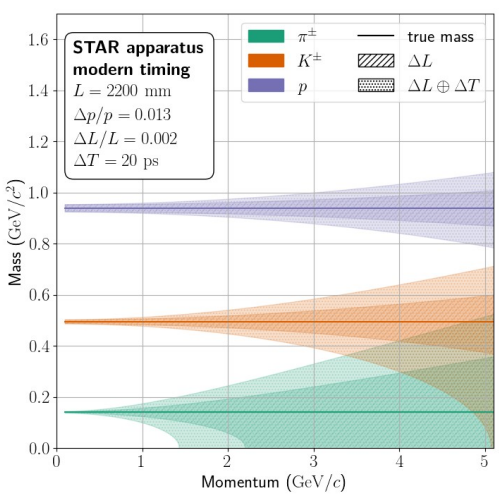
$$L = \sum_i^{N_{\text{hits}}} L_i = \sum_i^{N_{\text{hits}}} \frac{|z_{i+1} - z_i|}{|\tan\lambda_i|}\sqrt{1 + \tan^2\lambda_i}$$

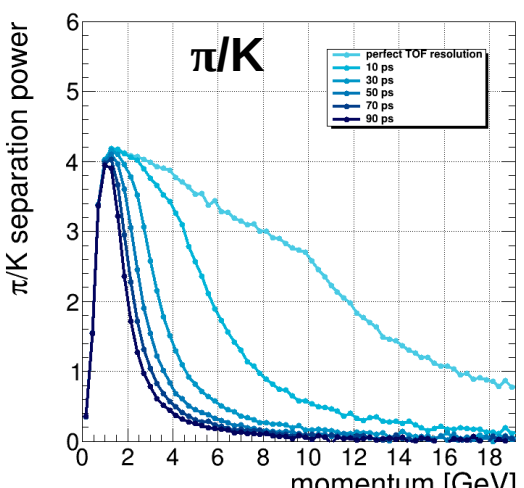
https://indico.desy.de/event/34916/contributions/147145/

Time of Flight

- Crucial: track length uncertainty may be a limiting factor to TOF performance
  - Example below: ΔT = 10 ps ~ ΔL = 3 mm
- p-value assessment of separation power includes outliers and gives more conservative estimate at low momenta (for details see backup)
- Still missing: digitizer; e.g. effect of hit energy deposition on hit timing



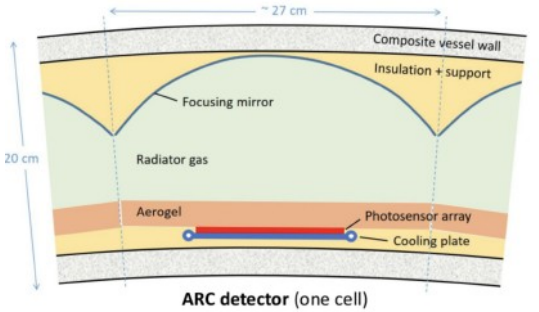




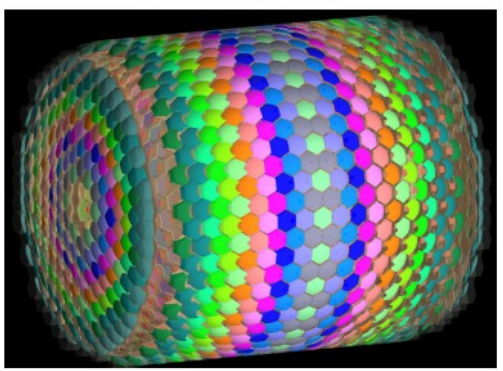
https://indico.desy.de/event/34916/contributions/147145/

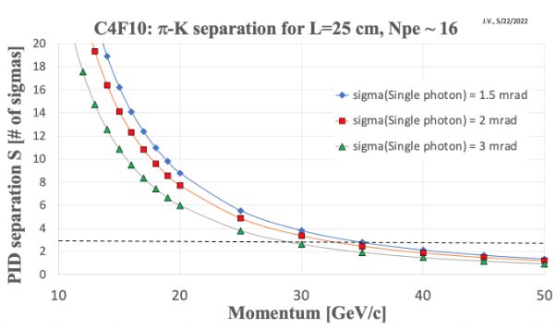
Ring Imaging Cherenkov Detectors

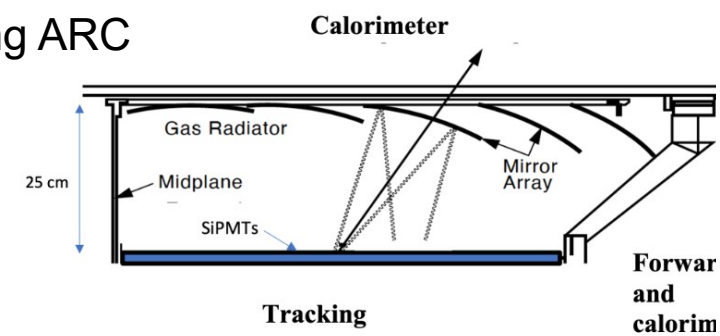
- 2 hardware proposals, aiming at PID up to 50 GeV with compact barrel+endcap RICH
- RICH for e.g. SiD, single phase
  - work ongoing on hardware and geometry
- ARC for CLD, with aerogel and gas
  - work ongoing on digitisation and reconstruction
  - allow for parametrised detector
  - provide CLD model including ARC

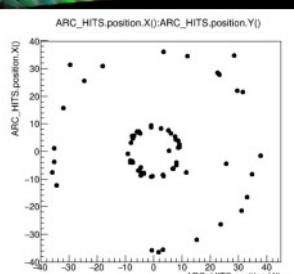


ARC detector (one cell)







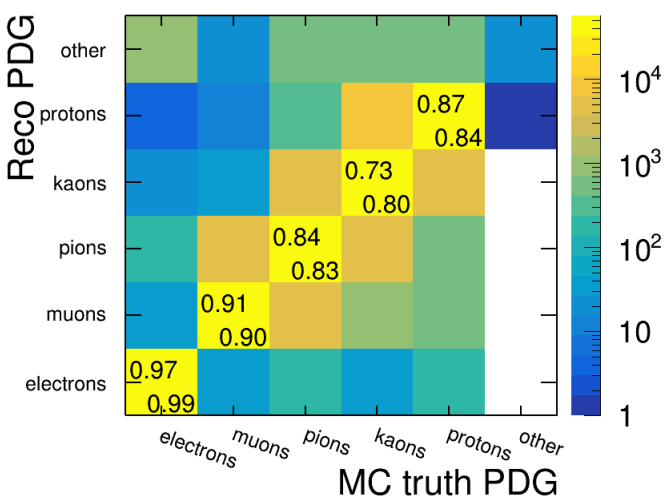


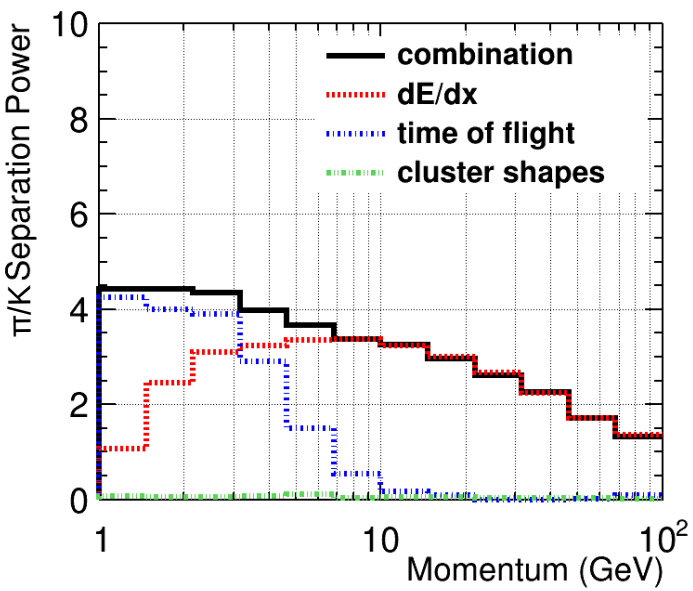
https://arxiv.org/abs/2203.07535

https://indico.cern.ch/event/1202105/timetable/#354-status-of-software-for-det

Comprehensive PID

- Modular approach to combined PID, both for the input observables and the training models
- Using PID observables from existing reconstruction, modules for these inputs as well as the training models to combine them
- Allows to optimise and compare different PID ‘settings’ in a detector or different detector with each other



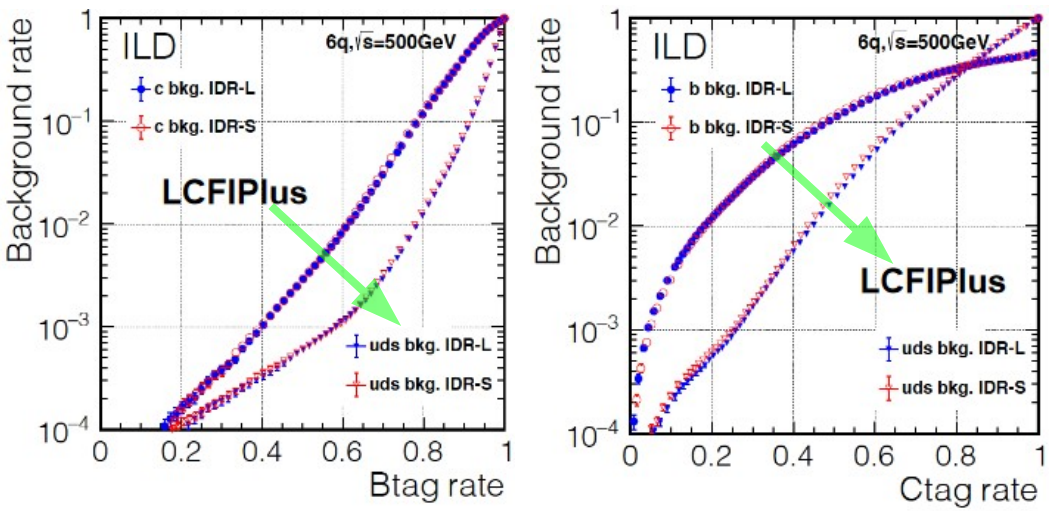


https://indico.cern.ch/event/1283129/#1-a-comprehensive-particle-id

LCFIPlus

- Well-established ‘standard’ tool
- Using mostly vertex jet constituent information via BDT for tagging
- Output: *b*, *c* or *other* tag
- No hadron PID yet

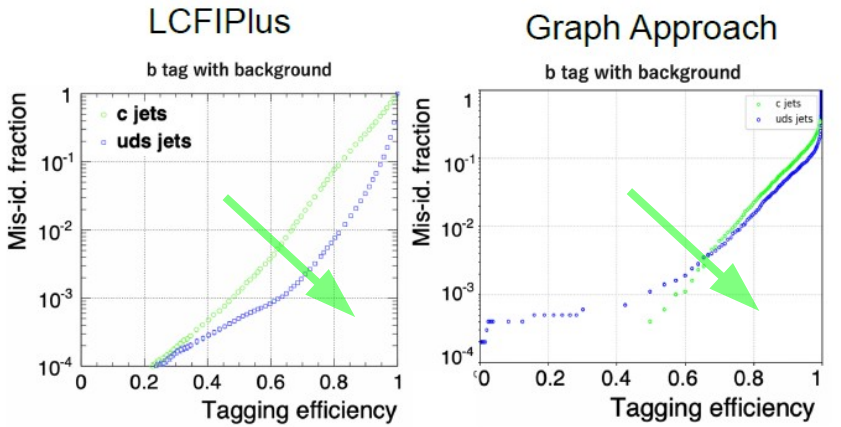
- For reference, arrow indicates in each tagging plot the ‘desired’ direction



https://arxiv.org/abs/2003.01116

Graph Attention Network

- Use Graph Neural Network (GNN) with ‘attention mechanism’
- With same input/output as LCFIPlus it shows improvement at high efficiency values
- Some fine-tuning still needed



Attention mechanism ... Learn the importance score for each weight
Take as a coefficient for update parameter.
Aimed by attention expressing whether tracks has the same vertex.

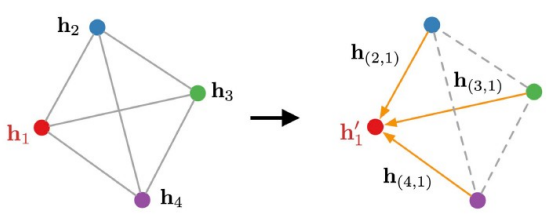


Fig. Graph Training

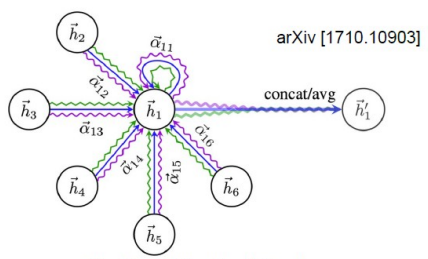
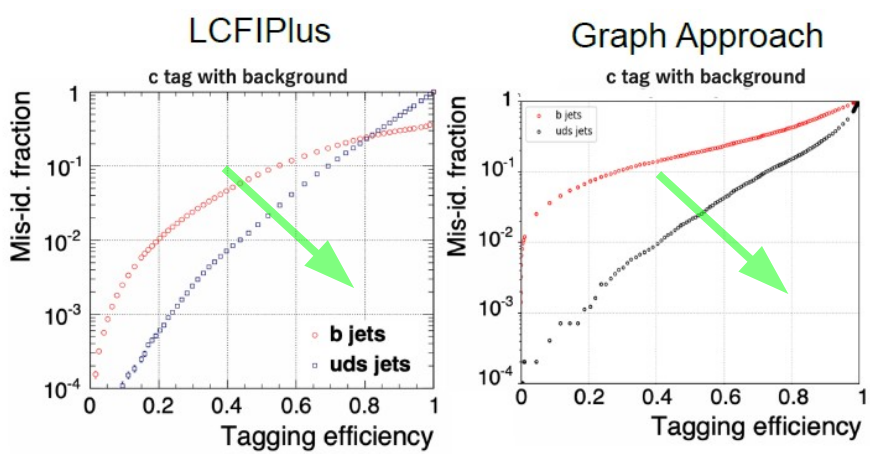


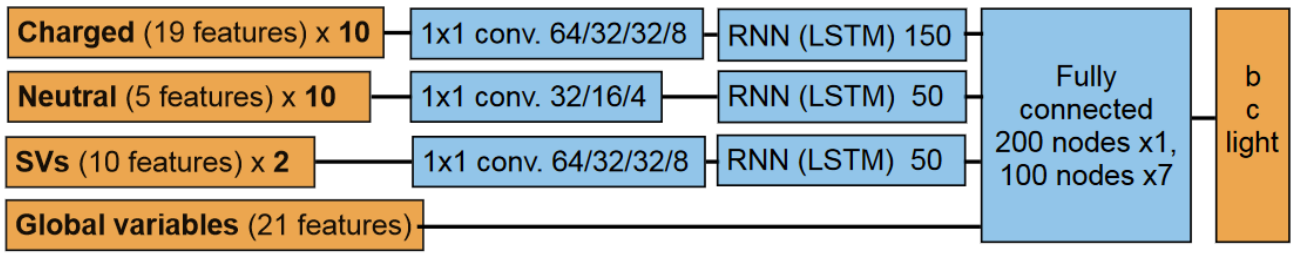
Fig. Graph Attention Network

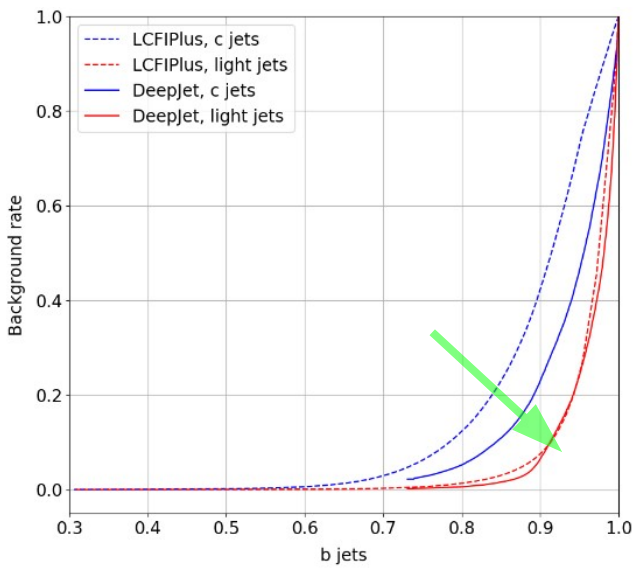


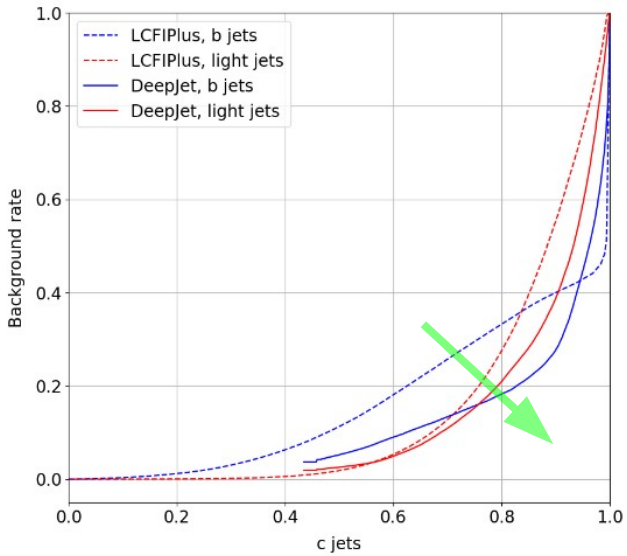
https://indico.slac.stanford.edu/event/7467/contributions/5948/, https://agenda.infn.it/event/34841/contributions/208239/

DeepJet and ParticleNet (ILD)

- Similar input to LCFIPlus, same output classes
- Applied to ILD MC data
- Based on CMS DeepJet
- Overall improvement wrt. LCFIPlus



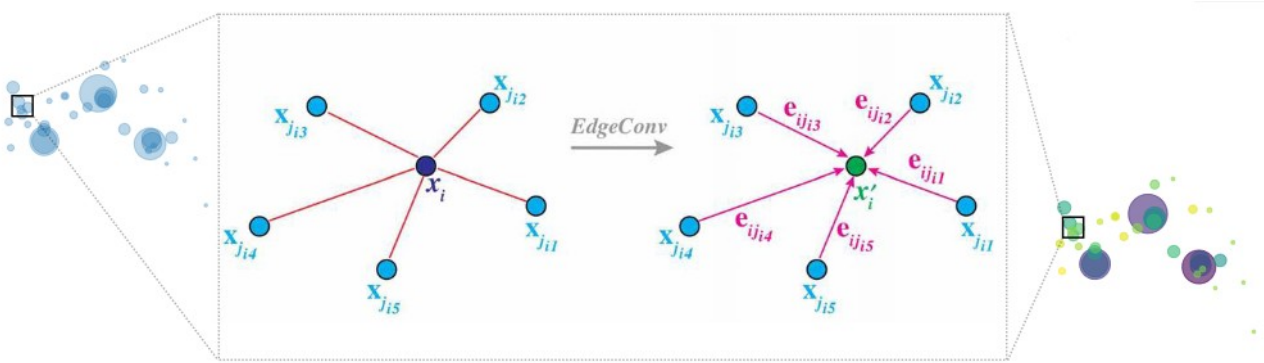


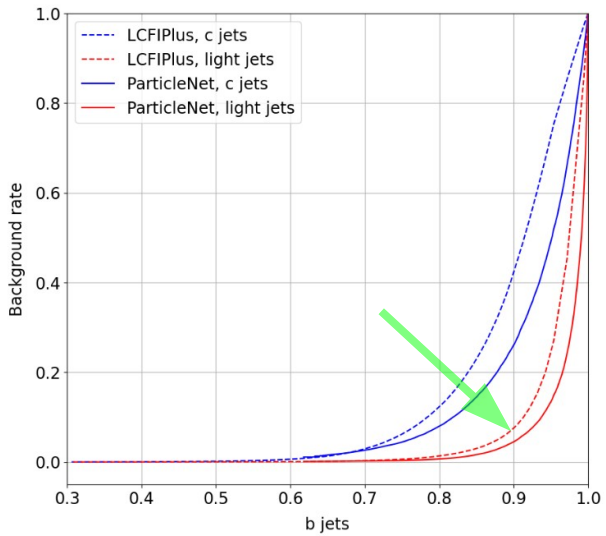


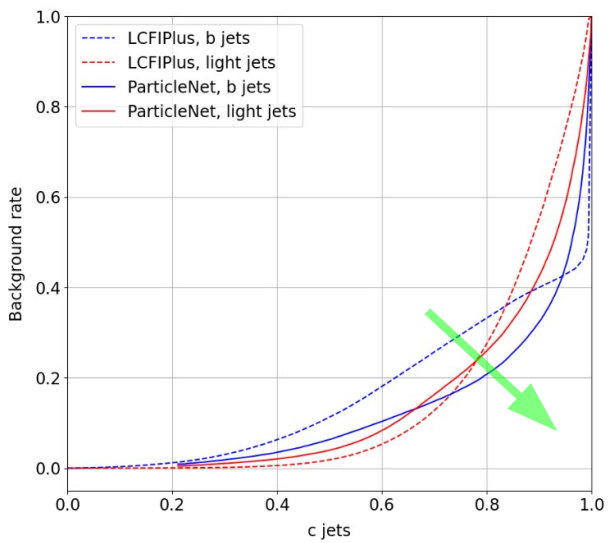
https://agenda.infn.it/event/34841/contributions/207748/

DeepJet and ParticleNet (ILD)

- Similar input to LCFIPlus, same output classes
- Applied to ILD MC data
- Based on ParticleNet GNN
- Overall improvement wrt. LCFIPlus
- ParticleNet at a slight advantage, but optimisation is ongoing



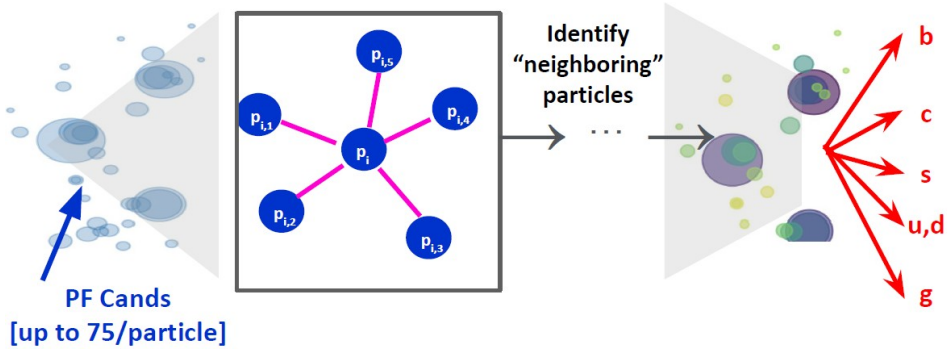


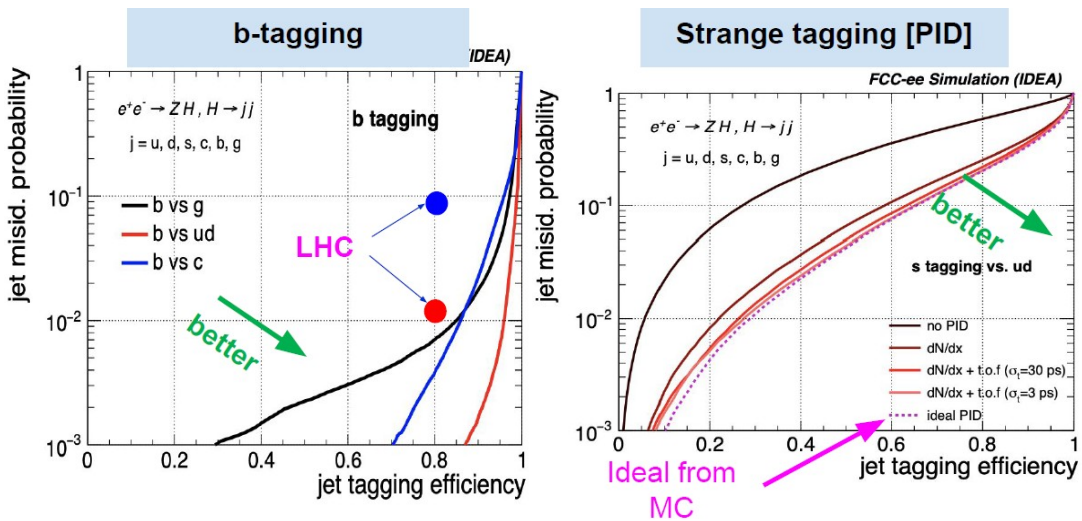


https://agenda.infn.it/event/34841/contributions/207748/

ParticleNet (IDEA)

- Again, ParticleNet based on GNN
- Applied to IDEA (Delphes)
- Making use of full detector performance incl. PID
- Output: b, c, s, u/d, g
- Excellent performance, allows for strange tagging
- Used for extensive Higgs branching ratio study

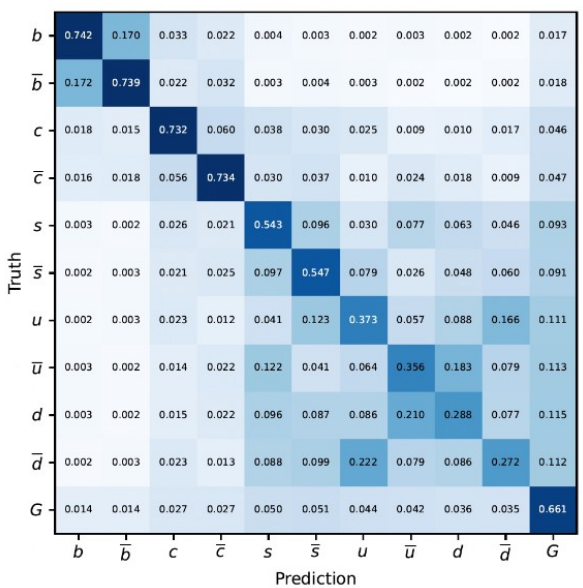


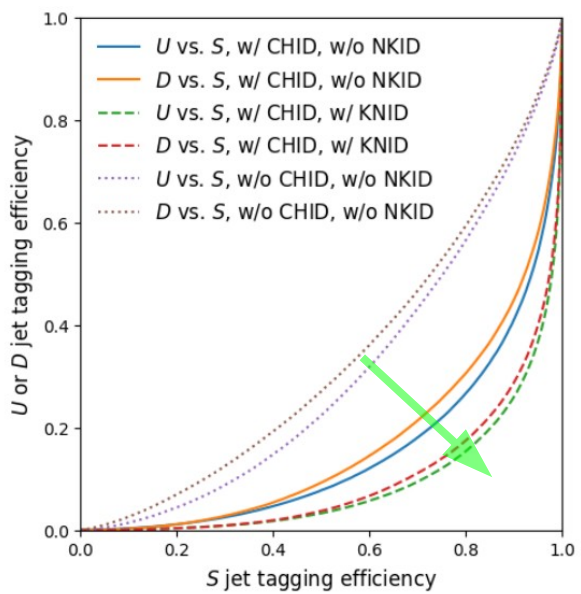


https://indico.cern.ch/event/1283129/#6-particlenet-ml-tagger, EPJ C 82 646 (2022)

ParticleNet (CEPC Baseline)

- CEPC baseline detector full sim and Arbor reco; incl. TPC & V^0 PID for kaons
- 11 output categories: 5 q, 5 $\overline{q}$, g; shown in confusion matrix
- Shows impact of kaonID on s vs. d and u

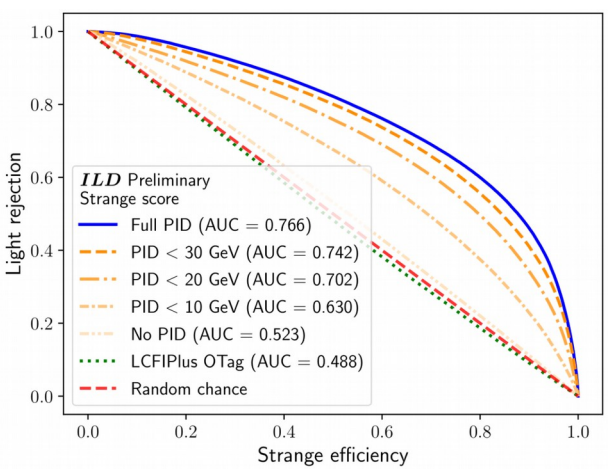


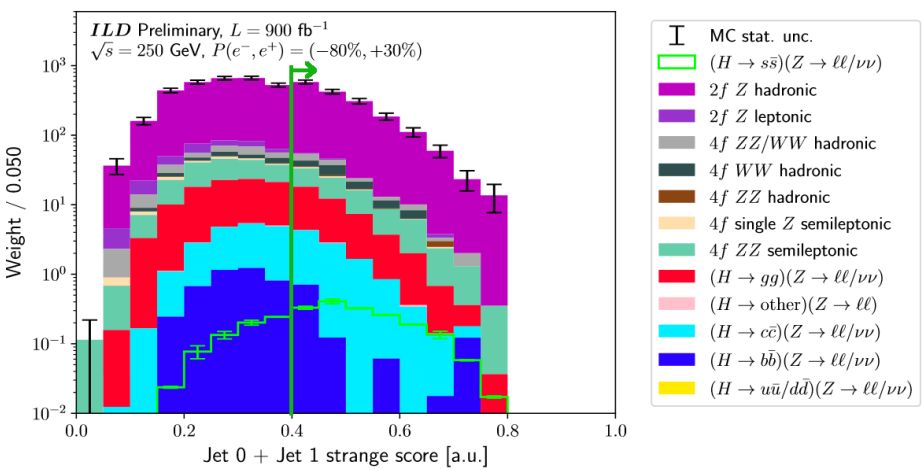


https://agenda.infn.it/event/34841/contributions/211981/

Self-made taggers

- Lack of readily available ‘new’ taggers has led to analysis-level taggers, with usage of LCFIPlus output and additional variables as input
- E.g. dedicated strange tagger, using c/b/o tag and PID in a separate BDT, analysis to determine strange Yukawa coupling

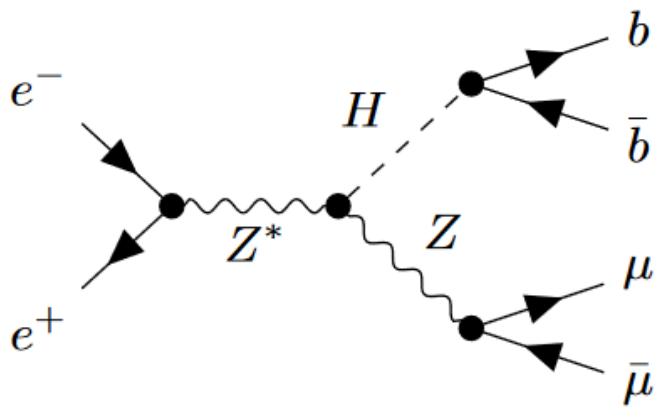




https://arxiv.org/abs/2203.07535

High-Level Reco Example

- Target here: hadronic Higgs
- Also applied to Higgs self-coupling in double H production, with 4 or 6 jets
- Difficult to distinguish hadronic H from hadronic Z background (e.g. ZHH vs. ZZH)
  - Neutrino correction
  - LeptonID
  - ErrorFlow kinematic fit

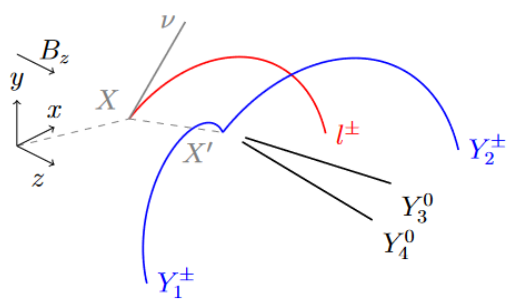


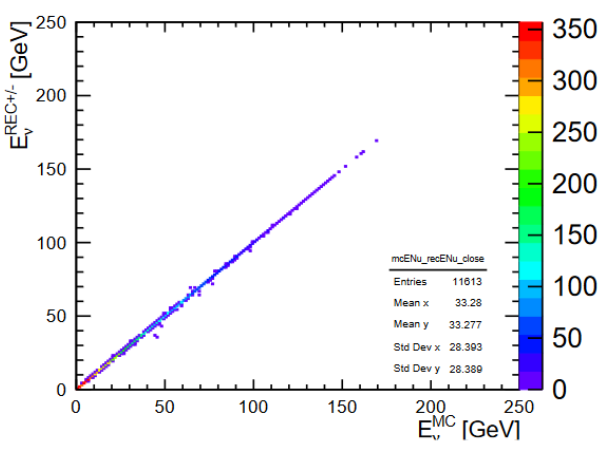
https://agenda.infn.it/event/34841/contributions/208287/

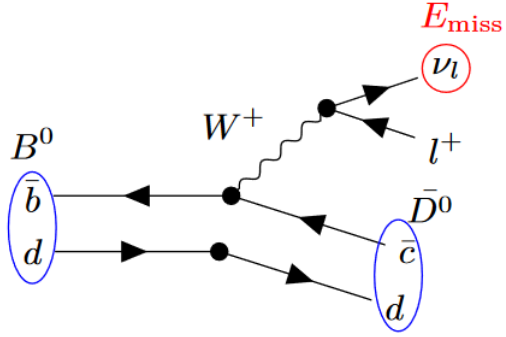
Neutrino Correction

- Semileptonic decays of b and c-hadrons have missing neutrino energy
- Traditionally: down-weigh event if lepton is seen in jet
- New: with constraints from e+e- collisions and finding* all consecutive vertex constituents, the neutrino momentum can be fully reconstructed*
  - * and identifying, at least the leptons
  - * up to a sign ambiguity

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



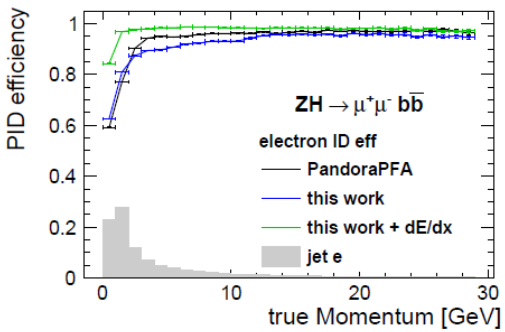


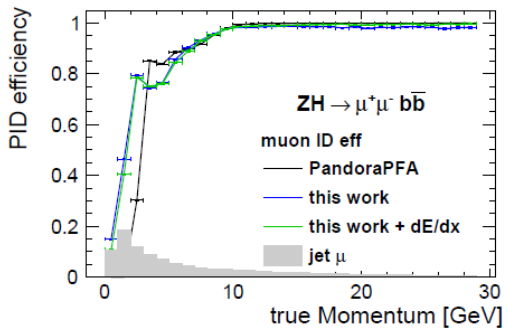


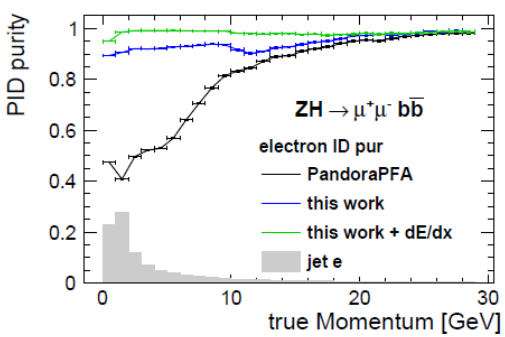
https://agenda.linearcollider.org/event/9823/#5-errorflow-jet-error-estimati

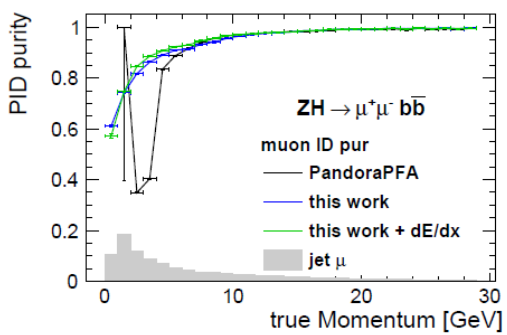
LeptonID

- Dedicated electron and muon ID to improve on ‘standard’ output from Pandora
- Put cluster shapes (~similar to Pandora) and dE/dx information into dedicated BDT
- Improves electron ID considerably, muon ID a bit









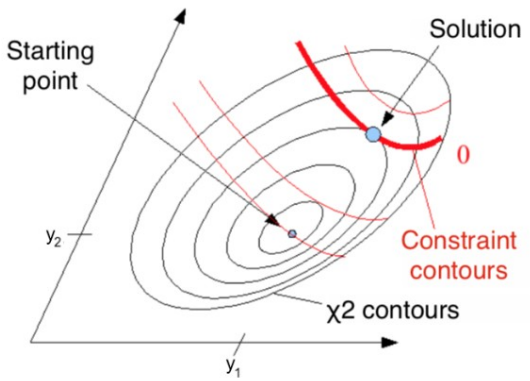
https://indico.cern.ch/event/1283129/#13-lepton-id-in-jets

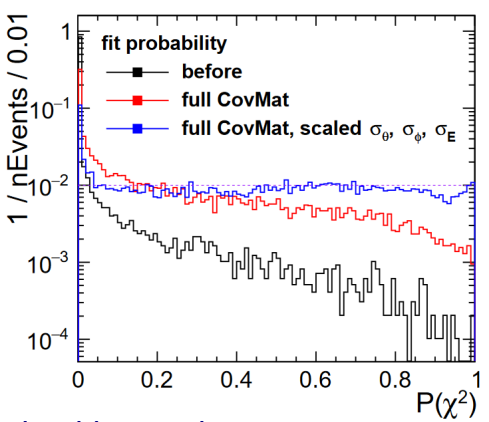
ErrorFlow Kinematic Fit

- Kinematic fit: re-assess measured properties of event within error bars (i.e. χ^2 fit) given general or analysis-specific contraints
  - e.g. re-evaluate jet energies, given the detector jet energy resolution, to satisfy $\Sigma E = 90$ GeV
  - done since LEP
- New: use ParticleFlow to determine uncertainties on individual objects (particles, jets, etc.) instead of generic detector uncertainties and re-weigh assigned measurement uncertainties to satisfy flat χ^2 → ErrorFlow
- Bonus: can cover the sign ambiguity in neutrino correction

χ^2-function to minimise:

$$L(y) = \Delta y^T \mathbf{V}(y)^{-1} \Delta y + 2 \sum_{k=1}^{m} \lambda_k f_k(a, y)$$

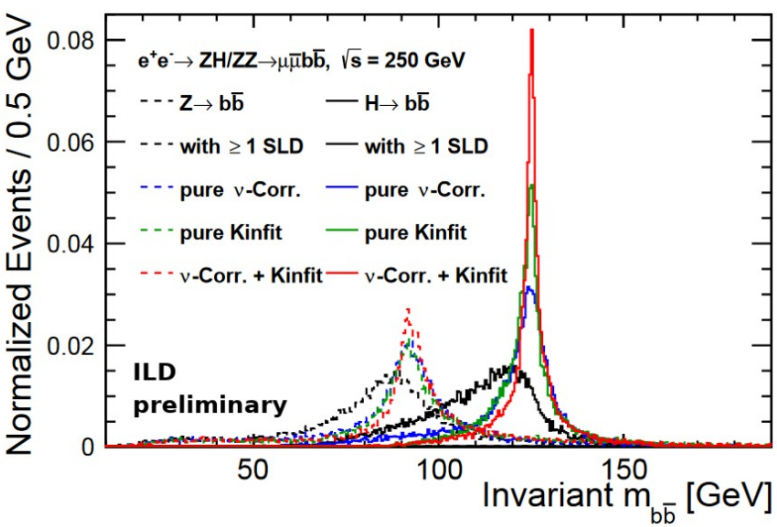




https://indico.cern.ch/event/1283129/#20-reconstructing-higgs-pair-p

Result

- Method reconstructs semi-leptonically decaying bosons with very high accuracy
- Allows to utilise great detector precision and pristine conditions at e+e- collider, in this case to separate Z from H peaks
- Example of combination of coherent contributions to central reconstruction tool box



https://indico.cern.ch/event/1283129/#20-reconstructing-higgs-pair-p

Conclusions

- Reconstruction is key to make the most physics out of the precious collisions
- Key4hep established as central tool, full sim + reco is on the way
- Several detectors still in iLCSoft, need personpower to eventually make full transition, make use of wrapper until then

- Various new ideas and implementation, in particular in PID and flavour tagging

  → make tools generic to use in Key4hep

  → coherently compare flavour taggers once development is ‘finished’ and aim to combine and optimise

Backup

The Future-Higgs-Factory Landscape

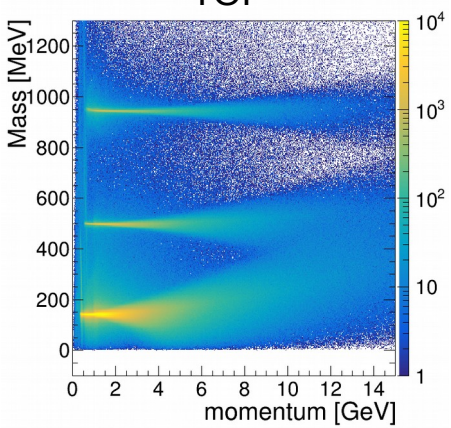
	CLIC	ILC	FCC-ee	CEPC	C^3	HALHF
• Full-Si	CLICdp	SiD	CLD	FST	SiD	
• TPC		ILD		Baseline		ILD
• DC, DR Calo			IDEA	IDEA		
• *novel approach*			ALLEGRO	4th Concept		

- Most new work in Key4hep (native EDM4hep/Gaudi)
- Still contributions in iLCSoft (native LCIO/Marlin), available in Gaudi via wrapper
- Some work being considered ‘analysis’ rather than ‘reconstruction’ happening in separate software, e.g. root macros

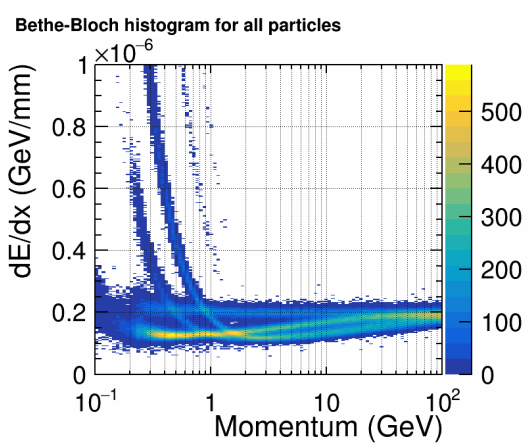
TOF low-mom level-out

- Compare to e.g. dE/dx: dE/dx takes ~ best 70% of hits, TOF needs accurate start-to-end measurement → small irregularities are cut out at dE/dx, but can screw up the TOF measurement
- B. Dudar has improved on this in ILD a lot, but still a 'background contamination' remains, O(%) or more
- For separation power use 'p-value assessment', i.e. not Gauss fit (TOF mass is not sufficiently Gaussian), but sep. pow. of two Gaussians which have the same overlap as the bands under questions (e.g. pions and kaons) → leads to levelling out of separation power at low momenta at ~4 corresponding to %-contamination
- This is just a different (more honest) view of the same data
- The relevant S > 3 area is still the same

TOF



dE/dx



- Find cut with mis-ID = 1 - efficiency = p-value → find Gaussian quantile → compute Z = 2 · quantile of standard Gauss