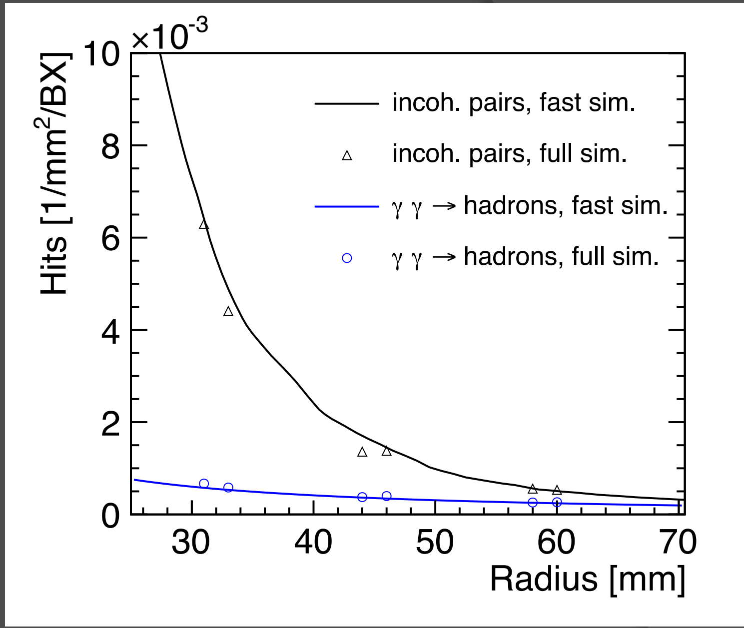
Tomas Lastovicka (Czech Academy of Sciences)
Katja Seidel (MPI Munich)
Jan Strube (CERN)

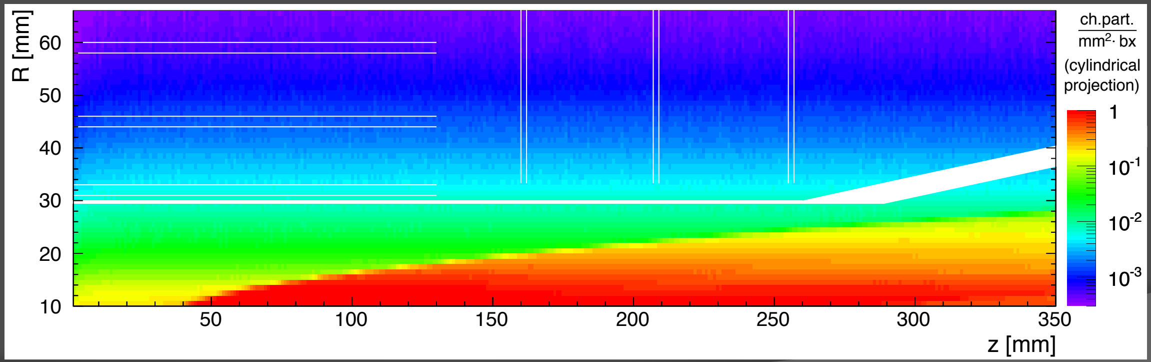
FLAVOUR TAGGING AT CLIC

Vertex detector design considerations

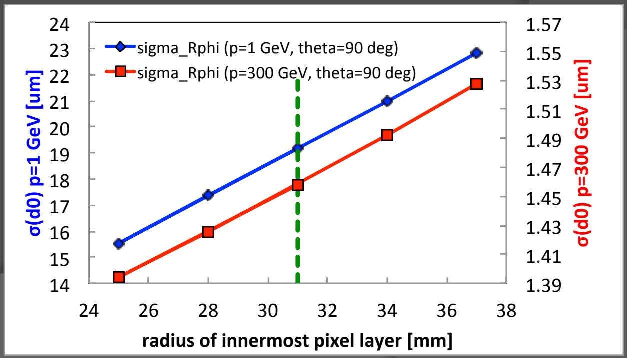
- Occupancies in the inner layers vs. single point resolution
  - → Field strength
- Number of hits vs. material budget



Barrel occupancies in CLIC_ILD



Direct hits from incoheren pairs, CLIC_ILD



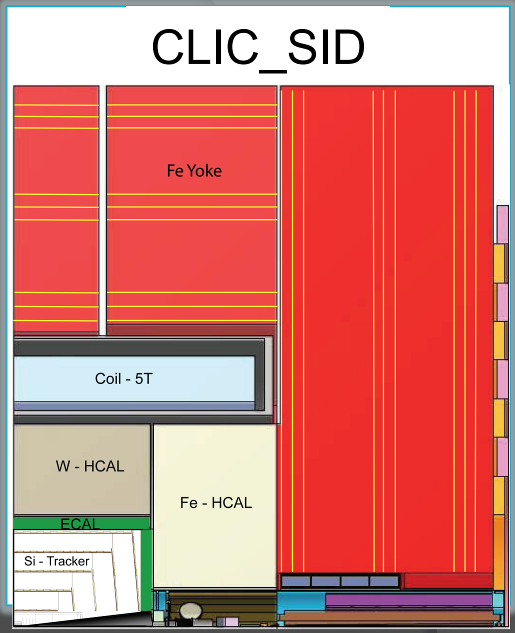
Tracking Detectors at CLIC

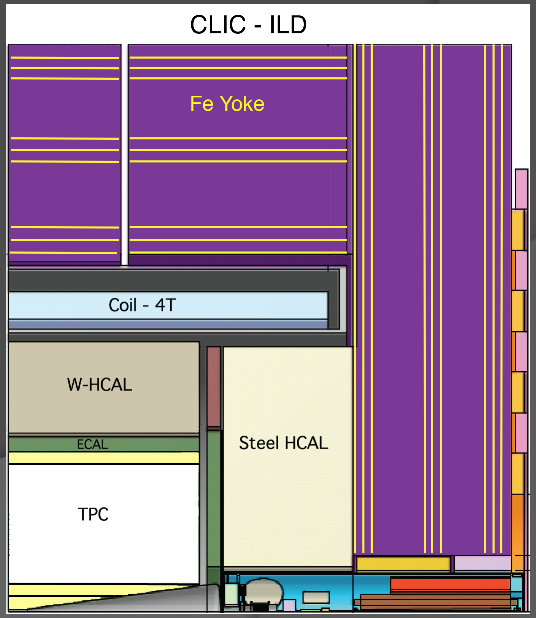
CLIC_SID

- Silicon Tracker
- 5 layers pixel
- 5 layers strips

CLIC_ILD

- 3 double layers pixel
- TPC + silicon envelope



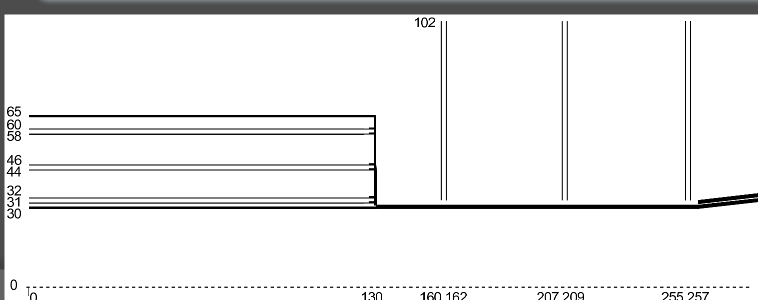


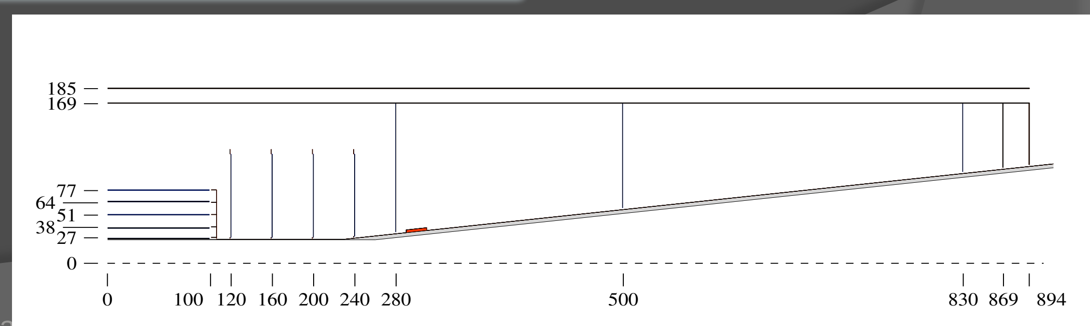
The Vertex Detectors

Table 4.1: Main parameters of the CLIC_ILD and CLIC_SiD vertex region layouts.

	CLIC_ILD	CLIC_SiD
Central beam pipe	Beryllium	
	$R_i = 29.4$ mm	$R_i = 24.5$ mm
	$d = 0.6$ mm	$d = 0.5$ mm
Barrel region	3 double layers	5 single layers
	$\|z\| < 130$ mm	$\|z\| < 98.5$ mm
	$R_i = 31,\ 44,\ 58$ mm	$R_i = 27,\ 38,\ 51,\ 64,\ 77$ mm
Forward region	3 double layers	7 single layers
	$z = 160, 207, 255$ mm	$z = 120, 160, 200, 240,$ 280, 500, 830 mm
Sensors	20 µm × 20 µm, $\sigma_{sp} \approx$3 µm	
	$X/X_0 = 0.18\%$ per double layer	$X/X_0 = 0.11\%$ per single layer
Surface area	0.736 m^2	1.103 m^2
Number of channels	1.84×10^9	2.76×10^9

For 500 GeV, lower background rates allowed to move inner pixel layer to 25 mm





Flavour Tagging

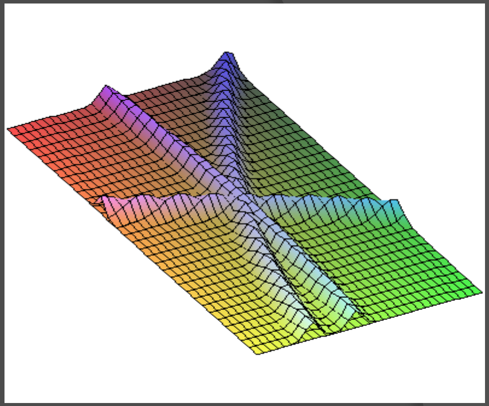
LCFI package:

- ZVTOP topological vertex finder
- Jet-based tagging
  - p_T - corrected mass
  - Impact parameter resolution
- Some neural network implementation for the tagging
  - 16 input variables per jet

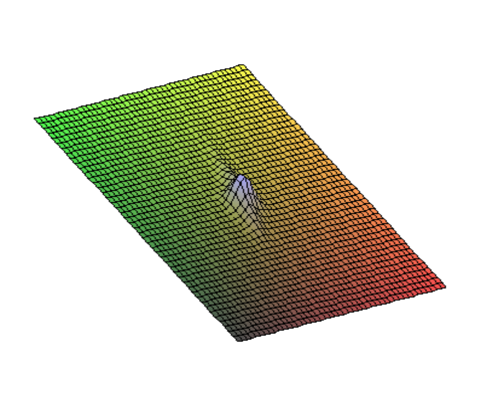
The Algorithm – ZvTop

D. Jackson, NIM A388:247-253, 1997

- Topological Vertex Finding
- Looks for overlap of Track probabilities in the Jet
- Resolves ambiguities with a resolution criterion
- Tracks are fitted to the point of highest overlap

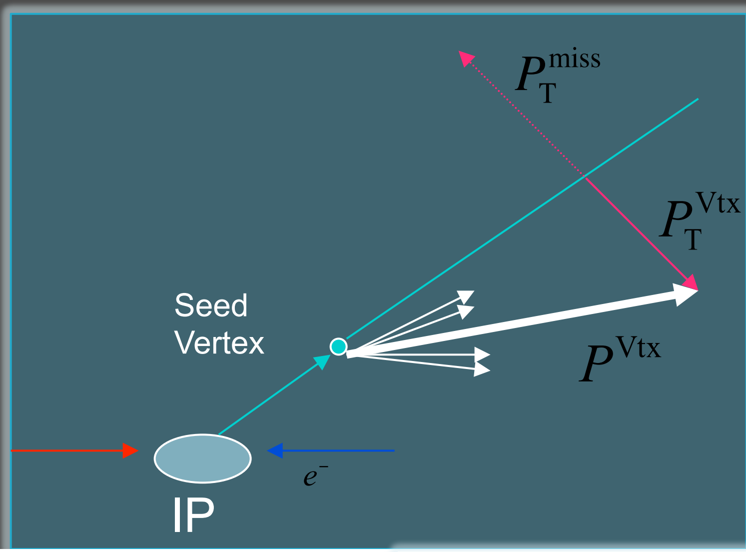


Tracks are assigned Gaussian Tubes

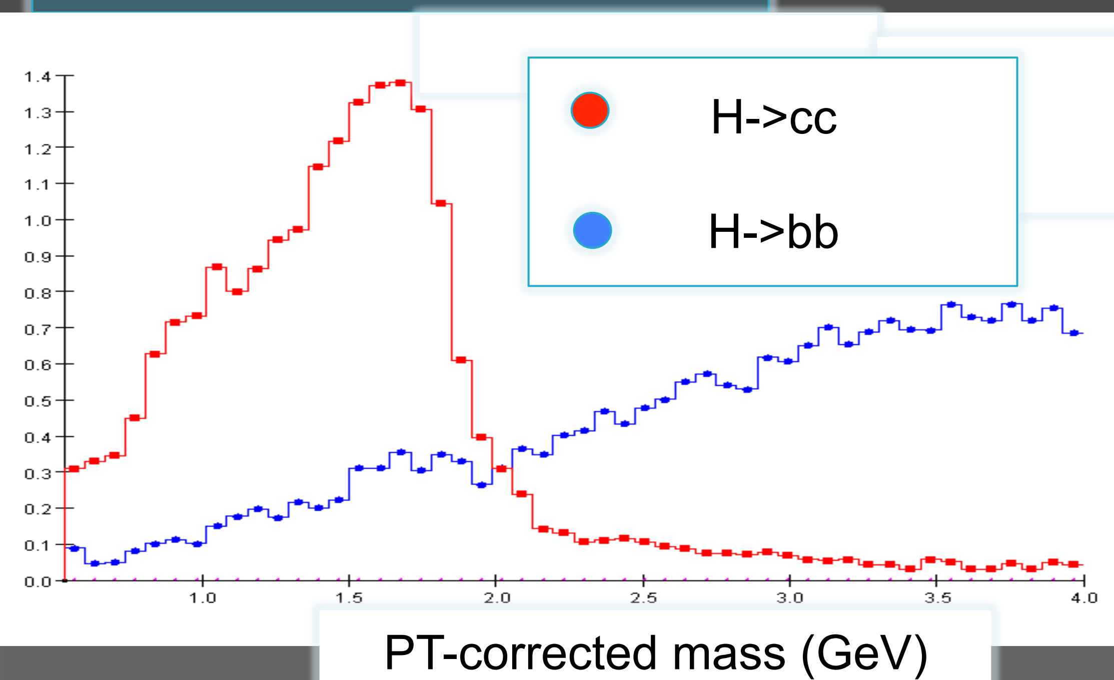


The maximum overlap is calculated

Correcting the vertex mass for neutral particles



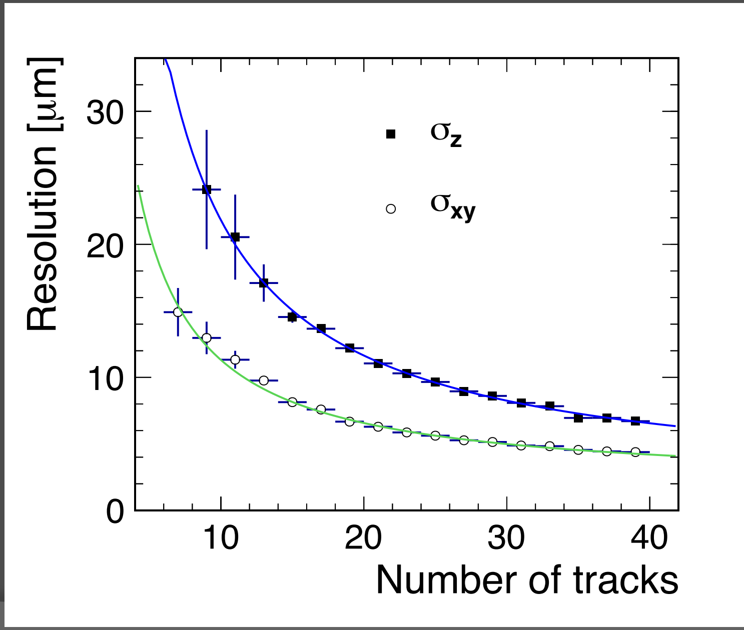
$$m_{p_T} = \sqrt{\left(\sum_{\text{tracks}} m_\pi\right)^2 + |p_T|^2} + |p_T|$$



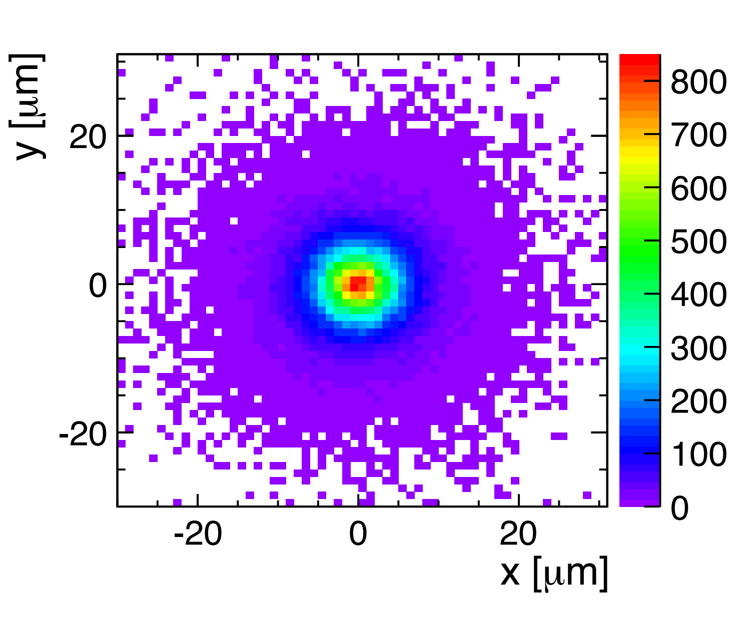
- The P_T-corrected Mass already by itself gives quite reasonable discrimination
- But one can see the contamination of B events in the D sample

Primary vertex resolution (CLIC_SID)

- LCFI doesn't use a primary vertex fit
  - But easier to get resolution, because of known position
- Use existing vertex fitter for primary vertex reconstruction
- Fit all tracks (modulo track quality) in the event to a common vertex



x-y view of primary vertices with more than 20 tracks



Detector layout cross-check post-CDR very preliminary

Use the 500 GeV ttbar events w/ different reco

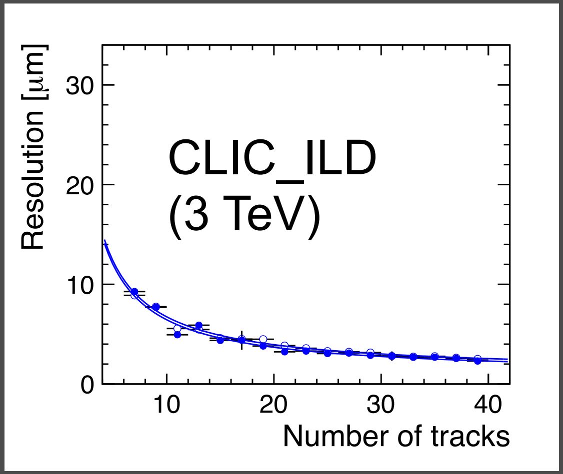
- Simulated and reconstructed w/ background

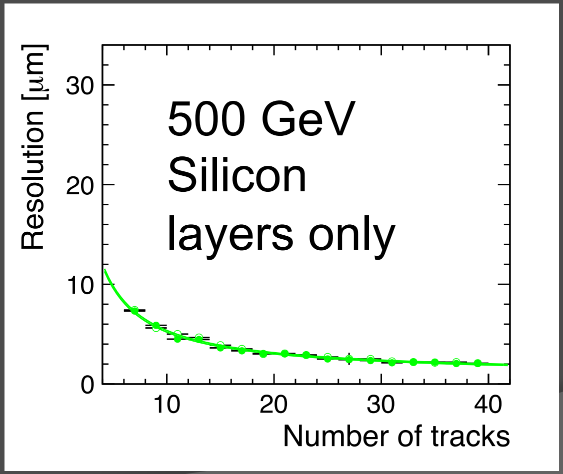
- Production (Si layers only for IP resolution)
- Full tracking for IP resolution
- Change rphi resolution parameter
- Simulate and reconstruct them in the CLIC_ILD (3TeV) detector

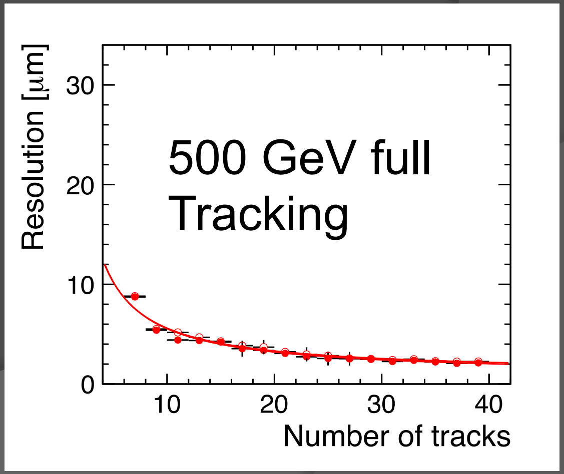
Use existing ttbar flavour tag nets for the analyses of the samples

Primary vertex resolutions

- Clear difference between primary vertex resolution for 3 TeV detector and 500 GeV detectors
- Little difference between the different tracking options in the same detector

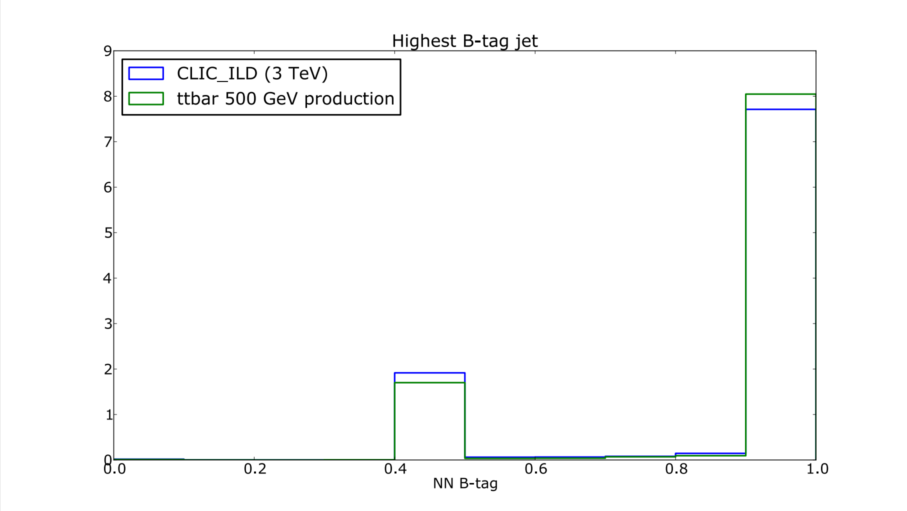


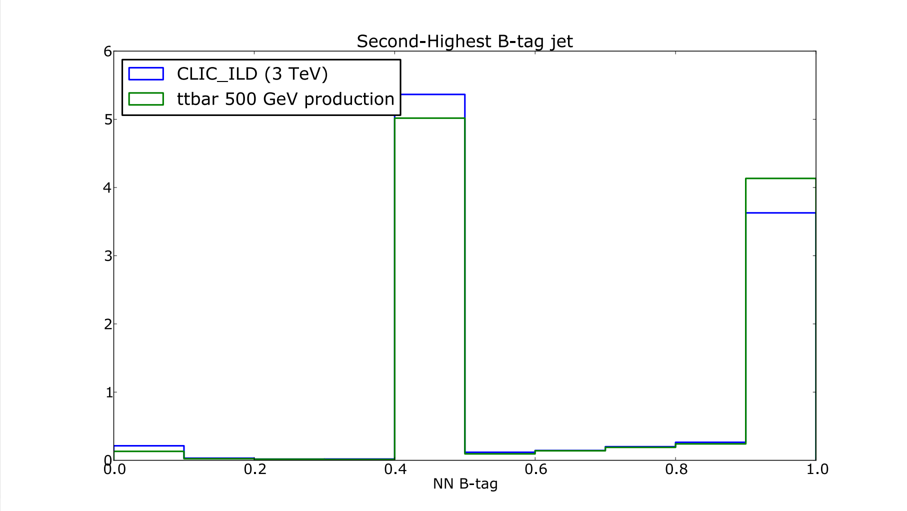




B-tagging performance

- Different geometries lead to different b-tagging performance
- Good cross-check of physics-driven requirements for layout choices

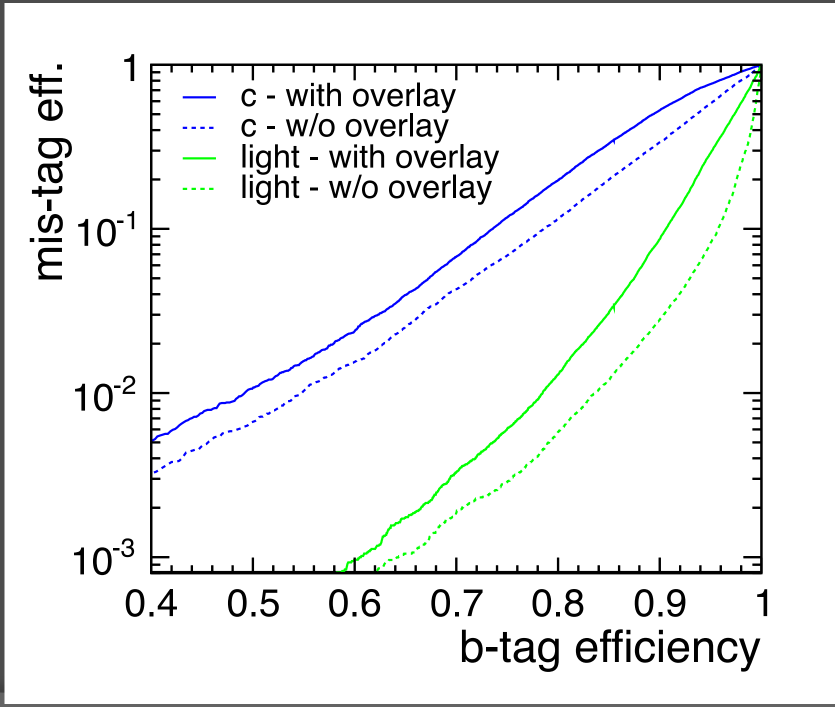


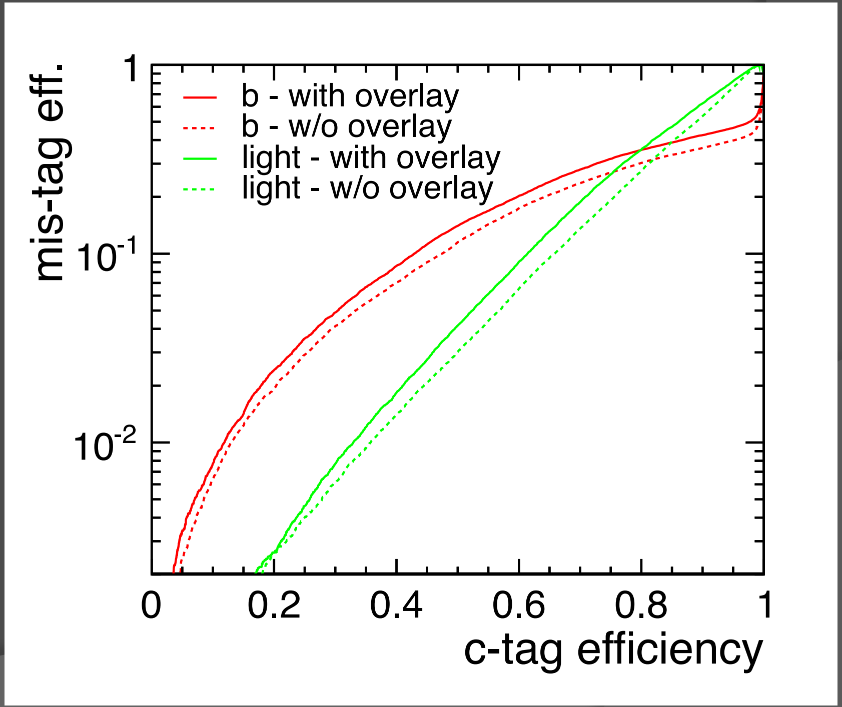


Flavour Tagging in the CDR

Light Higgs decays to bottom and charm

- Mean energy of Jets 130 GeV
- Using FastNN for training
- Additional track-based variables used in additional step
- b and c(!) tagging



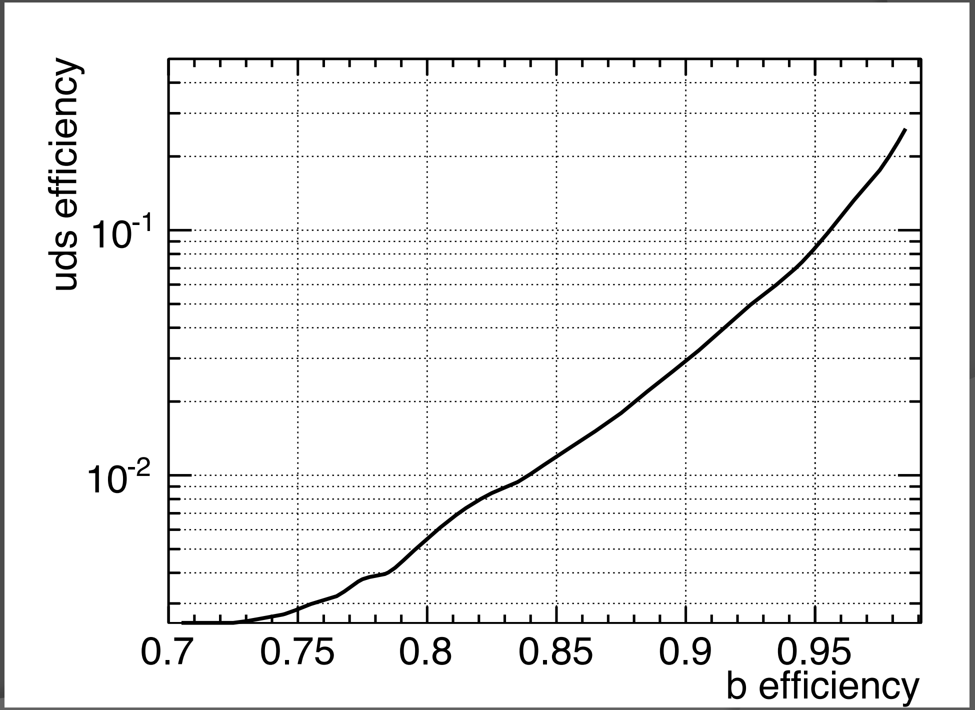


Heavy Higgs

- 30% of secondary tracks from beyond innermost vtx layer
- 20-30% of energetic b jets and 90 % of light quark jets have no secondary vertex

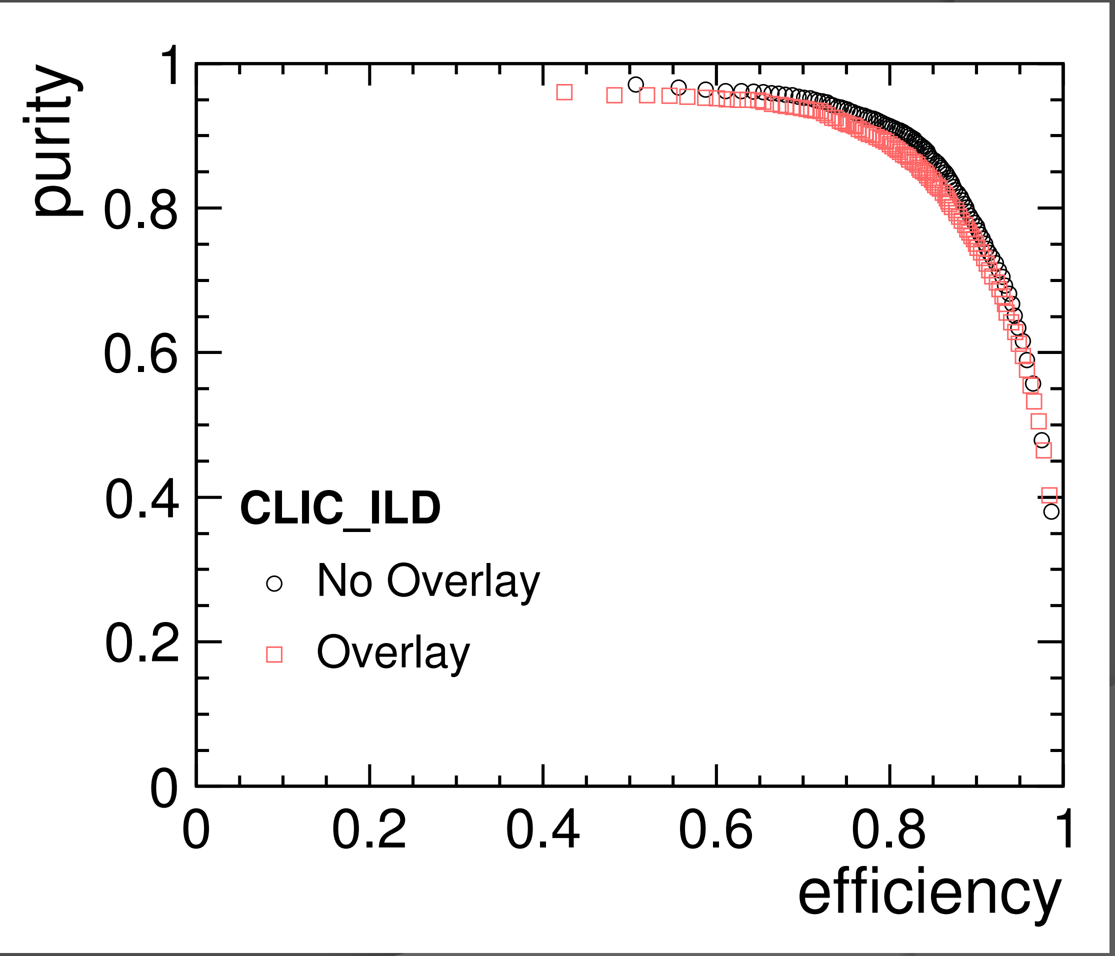
TMVA for training of the networks

Augmented with track-based variables



Top pairs at 500 GeV

- 6 jets, Anti-k_t, R=1.3
- Standard variables as defined by default LCFI
- TMVA for training



Lessons Learned during the CDR exercise

Evaluation of flavour tagging performance

- For simple topologies: can use “MC truth”
- For complex final states:
  - Training on simple topologies no good
  - Truth matching needed for training
    - H → bb: 5% inefficiency
    - Neutralinos decaying to Higgs: 20% inefficiency
    - Top pairs @ 500 GeV: 40% inefficiency

→ signal and background in complex final states need to be carefully defined

→ Truth matching inefficiencies lead to reduced number of events available for training

Preservation of expertise

- 3 very different physics analyses in the CDR
- 3 different levels of expertise
- 2 experts augmented the existing functionality, 1 used more or less default settings

→ Some lessons learned might be general enough to be fed back

Summary

- LCFI flavour tagging package was used successfully for b-tagging in CLIC CDR benchmarking studies
- The performance of the package for individual analyses is sufficient
- For large-scale studies and comparison, it would be desirable to contain the beginning fragmentation
- Options to involve a more direct feedback from LCFI for physics studies should be investigated