University Of Milano-Bicocca

Physics Department

Master's degree in Physics

MASTER THESIS

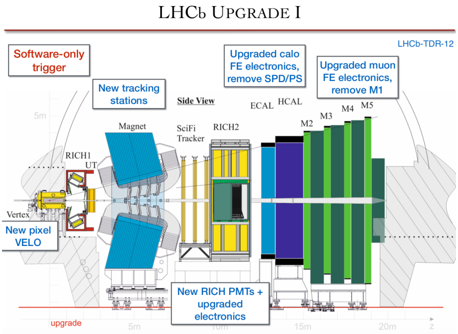
Development and Applications of Graph Neural Networks for the Trigger of the LHCb Experiment

Advisor:
Prof. Marta Calvi

Co-Advisor:
Dr. Julián García Pardiñas

Candidate:
Martina Mozzanica
Student ID: **869374**

LHCb (Large Hadron Collider beauty experiment)

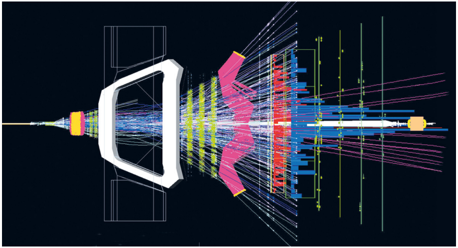


- The LHCb experiment is one of the four main experiments located along the **LHC ring at CERN**
- The LHCb detector is a single-arm spectrometer instrumented in the forward region

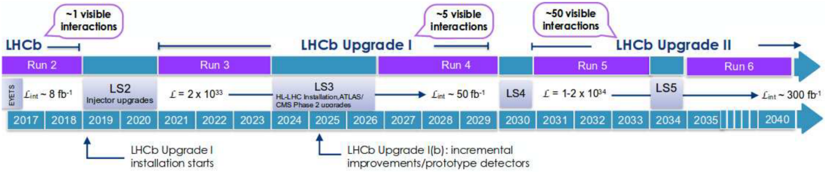
LHCb is a dedicated **b and c-physics precision experiment** for new physics beyond the Standard Model through the study of both very rare decays and **CP violation** in charm and beauty hadrons, aiming to better understand the asymmetry between matter and antimatter.

Upgrade at LHCb

The LHCb experiment has completed Upgrade I and is preparing for Upgrade II in a decade's time



$\Longrightarrow$ increasing of the instantaneous luminosity by five and ten times leading to unprecedented challenges for the trigger system due to an increase in signal and background rates



DFEI

- The current algorithms, used to make the trigger efficiently identify signals, would be too slow to deal with the high level of particle combinatorics.
- Event size will become too large to allow the storage of the full event information
- One expects to have multiple signals per event, while the approach until now is focused on the study of individual signals.

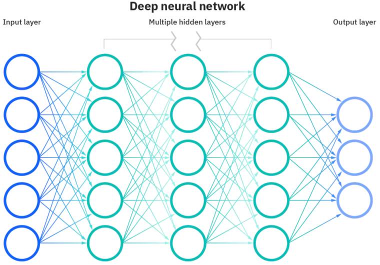
To address this issues:

LHCb Deep Full Event Interpretation (DFEI) project

- aims to design a deep full event interpretation algorithm, constructed as a sequence of *Graph Neural Networks*
- the algorithm is structured trough a chain of modules, which filter away part of the event.
- it uses *deep neural networks* to process information from all the stable particles in each event

GOAL: reconstruct b- and c-hadron decay chains while providing a good level of separation against all the other particles from the rest of the event

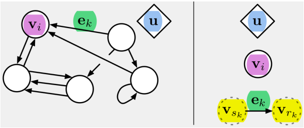
Graph Neural Networks



<u>def.</u> *Graph Neural Networks are neural networks that operate on graphs*

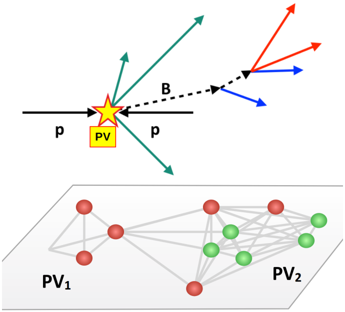
- A set of objects (*nodes or vertices*), and the connections between them (*edges or link*) are expressed as a *graph*.
- The global attributes for the graph block is u while s_k and r_k indicate the sender and receiver nodes indices for an edge k.

- Deep Neural networks include an input layer, one or more hidden layers, and an output layer. The layers are composed by neurons.
- Each neuron connects to another and has a weight and a threshold. If the neuron output is above the threshold value, that neuron is activated, sending data to the next layer of the network.



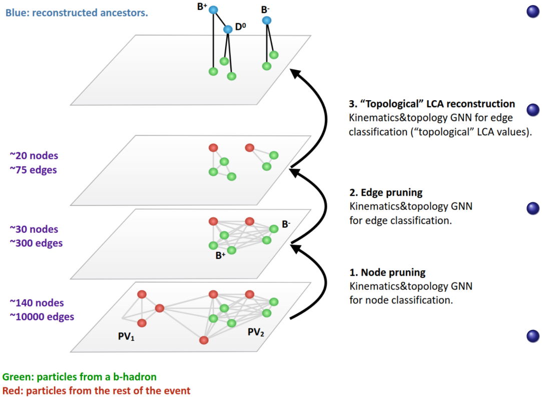
DFEI Algorithm

- **Input:** reconstructed stable particles
- **Basic Deep-Learning Unit:** Graph Neural Networks (GNN)
- **Algorithm Structure:** chain of modules with increasing "granularity", each of them filtering away part of the event
- **Graph:** all the information in each event
- **Nodes:** final-state particles
    * **Signal**: particles from a b-hadron decay
    * **Background**: particles from the rest of the event
- **Edges:** relations between each possible pair of particles



PV1, PV2: different proton-proton primary vertices.

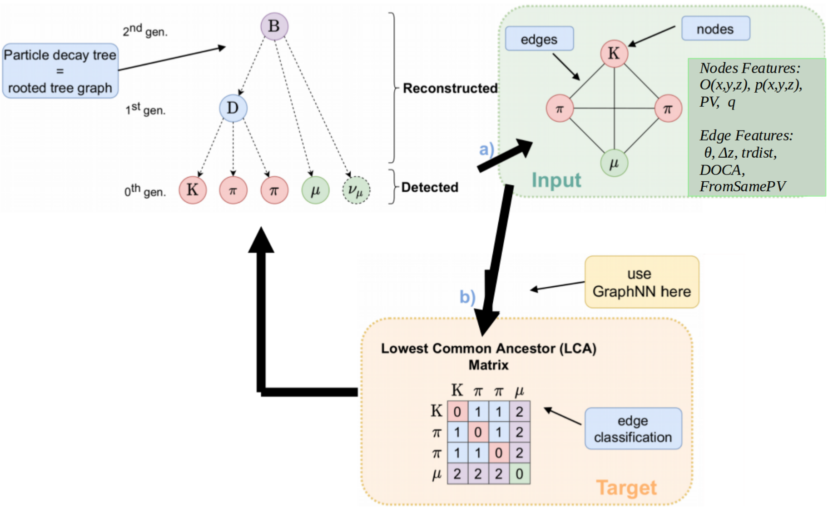
DFEI Algorithm



LCA = Lowest Common Ancestor

- **0. Edge Prefiltering:** preliminary reduces the number of edges (later explained)
- **1. node pruning:** removes particles that do not come from a signal heavy hadron
- **2. edge pruning:** removes links between particles that are topologically distant and produced by different signal heavy hadrons
- **3. topological LCA reconstruction:** uses a multi-class classification on edges to reconstruct the decay chain and produce a LCA matrix as output

LCA (Lowest Common Ancestor)



	K	π	π	μ
K	0	1	1	2
π	1	0	1	2
π	1	1	0	2
μ	2	2	2	0

Fast and Full simulations

- **Fast Simulation**:
  - based on PYTHIA8
  - emulated experimental effects of inefficiency, resolution and wrong PV association
  - fast to produce
- **Full Simulation**:
  - Geant4-based official simulation of LHCb events
  - all the interactions of the particles with the LHCb detectors and the detailed reconstruction processes are simulated
  - slower to produce, but more realistic
- both simulations are in Run3 conditions (multiple PVs) and done for this study

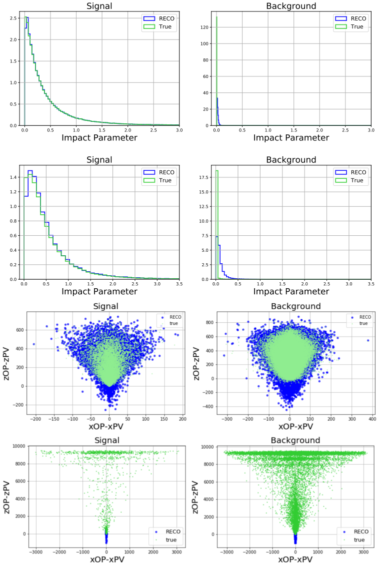
Number of:	Fast	Full
events	50k	10k
particles/event	145	99.9
b hadrons/event	1.5	1.7
particles from b hadrons/event	5.9	6.2

Fast

Full

Fast

Full



Truth Matching

In this thesis two types of event samples were used:

- **Inclusive sample**: of $b\bar{b}$ events that contains all types of b-hardon decays
- **Exclusive sample**: of events each one with a specific b-hadron decay mode

For this study it was needed to reconstruct the true decay tree of the b decay (in both cases) with the truth-matching algorithm.

	Branching Ratio	Fast Incl.	Full Incl.	Full Excl.	Fast Excl.
$B^+ \to \bar{D}^0[K^+\pi^-]e^+\nu_e$	$0.9 \cdot 10^{-4}$	123	19	-	-
$B^+ \to \bar{D}^0[K^+\pi^-]\mu^+\nu_\mu$	$0.9 \cdot 10^{-4}$	97	13	-	-
$B^0 \to D^-[K^+\pi^-\pi^-]e^+\nu_e$	$2.1 \cdot 10^{-3}$	40	6	-	-
$B^0 \to D^-[K^+\pi^-\pi^-]\mu^+\nu_\mu$	$2.1 \cdot 10^{-3}$	31	3	-	-
$B^0 \to K^+\pi^-$	$2.0 \cdot 10^{-5}$	14	7	359	4720
$B_s^0 \to \mu^+\mu^-$	$3.0 \cdot 10^{-9}$	0	0	498	-
$\Lambda_b^0 \to \Lambda_c^+[p\, K^-\pi^+]\pi^-$	$3.1 \cdot 10^{-4}$	3	7	342	1735
$B^0 \to K_s^0[\pi^+\pi^-]\pi^+\pi^-$	$1.8 \cdot 10^{-5}$	0	0	105	-
$B^0 \to K_0^*[K\pi]\mu^+\mu^-$	$9.3 \cdot 10^{-7}$	0	0	1162	4672
$B^0 \to K_0^*[K\pi]e^+e^-$	$1.0 \cdot 10^{-6}$	0	0	833	-
$B_s^0 \to D_s^-[K^-K^+\pi^-]\pi^+$	$1.6 \cdot 10^{-4}$	-	-	-	3937
$B^0 \to D^-[K^+\pi^-\pi^-]D^+[K^-\pi^+\pi^+]$	$1.9 \cdot 10^{-6}$	-	-	-	3372
$B^+ \to K^+K^-\pi^+$	$5.2 \cdot 10^{-6}$	-	-	-	4648
$B_s^0 \to J/\psi[\mu^+\mu^-]\,\phi[K^+K^-]$	$3.0 \cdot 10^{-5}$	-	-	-	4357

Table: Number of decays found with the Truth-Matching algorithm

This is a new designed algorithm (not the standard LHCb one) since DFEI uses a special format to represent the events for efficient computation.

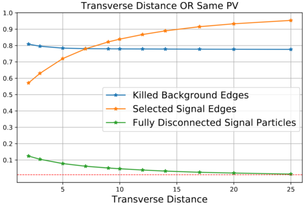
Edge pre filtering: 1D Cut

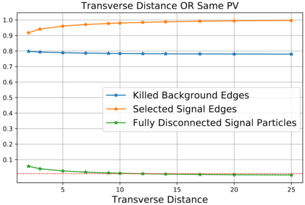
Edge Pre-filtering refers to a preliminary cut on the edges which will be implemented in the structure of the DFEI algorithm.

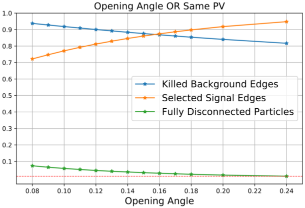
- *Selected Signal Edges (SSE)*: when a pair of particles coming from the same b-hadron decay overcome the threshold
- *Killed Background Edges (KBE)*: when a pair of particles that don't come from the same b-hadron decay do not overcome the threshold
- *Fully Disconnected Signal Particles (FDSP)*: when a particle coming from a b-hadron decay becomes fully disconnected from all the other particles of the same decay chain.

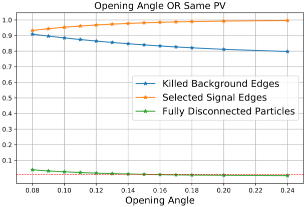
	Cut	FDSP (%)	KBE (%)	SSE (%)
Fast inclusive $b\bar{b}$	*Same PV*	19.4	77.3	46.2
	TrDist	20.9	59.7	49.8
	DOCA	21.2	38.4	47.1
	Theta	20.3	58.0	64.7
Full inclusive $b\bar{b}$	*Same PV*	18.4	77.9	77.4
	TrDist	19.8	55.2	73.1
	DOCA	18.1	17.4	78.2
	Theta	19.2	52.3	75.4

Edge pre filtering: 2D Cut









	Cut	FDSP (%)	KBE (%)	SSE (%)	Cut Value
Fast inclusive	*TrDist or Same PV*	1.0	77.8	96.6	25 mm
	Theta or Same PV	1.0	81.5	95.9	0.24 rad
Full inclusive	*TrDist or Same PV*	1.0	78.8	98.4	10 mm
	Theta or Same PV	1.0	84.6	97.6	0.14 rad

	Initial BKG Edges	Final BKG Edges
Fast	$1.8 \cdot 10^8$	$3.3 \cdot 10^7$
Full	$5.8 \cdot 10^7$	$8.9 \cdot 10^6$

Fast inclusive $b\bar{b}$ Simulation Performance

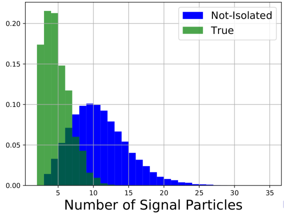
The edges preselection is now the nominal one applied.
DFEI was re-trained on the Fast simulation.

- *Perfect Event Reconstruction (PER)*: every signal in the event is perfectly reconstructed and all the background particles are removed.

- -

* *Perfect Signal Reconstruction (PSR)*: the reconstructed signal perfectly matches the true values.
* *Quasi Signal Reconstruction*: decays that have the same particles in the final state but has a different decay tree structure with respect to the target decay.
* *Not-Isolated Signal Reconstruction*: decays in which all the signal particles are contained in the reconstructed cluster but it also contains other (background) particles.
* *Partial Signal Reconstruction*: any other cases that do not fall into the above categories.

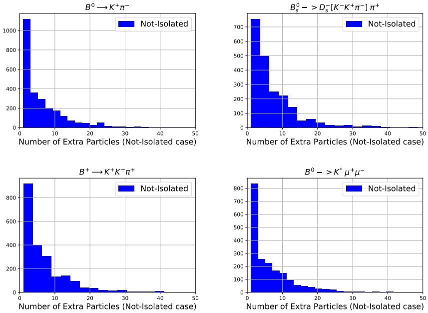
PER:	2.14 % ± 0.07%
PSR:	4.55% ± 0.08%
Quasi	5.92% ± 0.11%
Not-Isolated	76.04% ± 0.20%
Partial	13.43% ± 0.14%



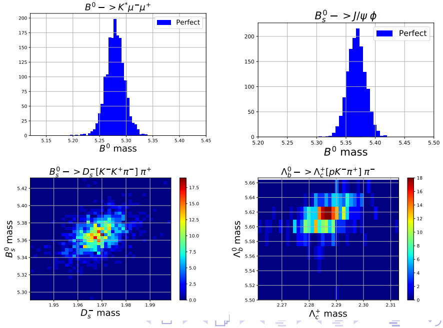
Performance of Fast Exclusive Samples

	PSR (%)	Quasi (%)	Not-Isolated (%)	Partial (%)
$B^0 \to K_0^*[K\pi]\mu^+\mu^-$	35.8 ± 0.7	19.2 ± 0.6	44.9 ± 0.7	0.0 ± 0.0
$B^0 \to K^+\pi^-$	38.0 ± 0.7	0.0 ± 0.0	54.7 ± 0.7	7.2 ± 0.4
$B_s^0 \to D_s^-[K^-K^+\pi^-]\,\pi^+$	32.8 ± 0.7	7.1 ± 0.4	53.7 ± 0.8	6.4 ± 0.4
$B^+ \to K^+K^-\pi^+$	35.7 ± 0.7	10.2 ± 0.4	46.4 ± 0.7	7.7 ± 0.4
$\Lambda_b^0 \to \Lambda_c^+[pK^-\pi^+]\,\pi^-$	21.7 ± 1.0	8.9 ± 0.7	36.8 ± 1.2	32.6 ± 1.1
$B_s^0 \to J/\psi[\mu^+\mu^-]\,\phi[K^+K^-]$	26.9 ± 0.6	20.5 ± 0.5	52.5 ± 0.6	0.0 ± 0.0

Extra Particles:



Mass Reconstruction:



Full Simulation Training of the LCA module

- The aim for the training is the **minimization of the loss (or cost) function** of which value usually gets lower for higher accuracies of the network's predictions.
- The training is done in a **supervised way**, using the ground truth present in the simulation.
- The true classes (LCA) are very **imbalanced**, which requires the uses of balancing methods in the loss.

Balanced Cross Entropy (BCE):

$$BCE = -\sum_{i=1}^{N} \alpha_i \log p_i$$

Balanced Focal Loss (BFL):

$$BFL = -\sum_{i=1}^{N} \alpha_i (1 - p_i)^{\gamma} \log p_i$$

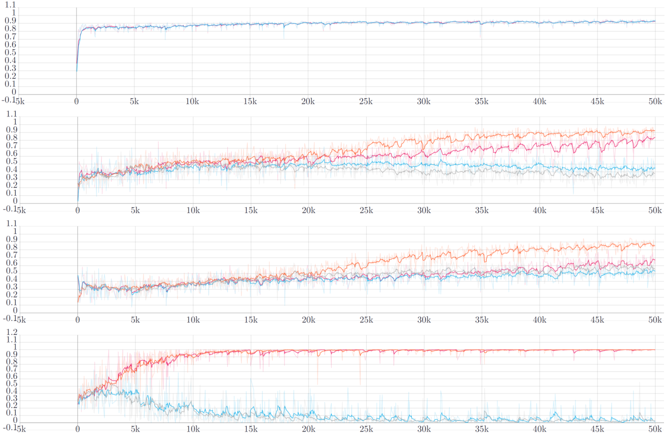
where $p_i = \sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$

x_i: element of the input vector
p_i: predicted probability (softmax)
N: total number of classes
α_i: weighting factor to each class
γ: focusing parameter

Training of the LCAs: BCE (Train Test), BFL (Train Test)

Different LCAs occur at **different rates** → accuracy for each LCA class.
LCA=0: most common, then LCA=2, LCA=1, and LCA=3: least common

`y axis: accuracy, x axis: iteration`



LCA=0

LCA=1

*

LCA=2

LCA=3

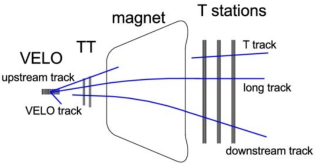
OVERTRAINING ! → more data are needed

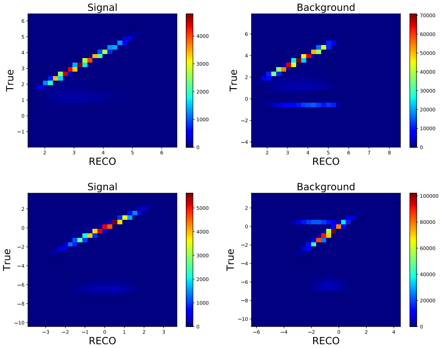
Conclusion

- This thesis expands a previous work on DFEI by using both the PYTHIA8-based framework, which had already been utilized, and the Geant4-based full LHCb simulation for the first time
- The development of a truth matching algorithm was an important step of the work since it has been used in most of the subsequent sections
- The pre selection incremented the accuracy with respect to the previous edge pre filtering cut
- The first detailed study of the DFEI performance on the Pythia-based simulation was performed for both the inclusive $b\bar{b}$ sample and exclusive decay modes samples.
- Even though the accuracy curves show how the focal loss performs better than the cross-entropy, the training was not successful due to the scarcity of data for LCA $=$ 1, 2, 3. Obtaining more data is the next step to properly train the DFEI algorithm on the full simulation.

Thank you

Cut of the Full Simulation





1D Edge prefiltering cut

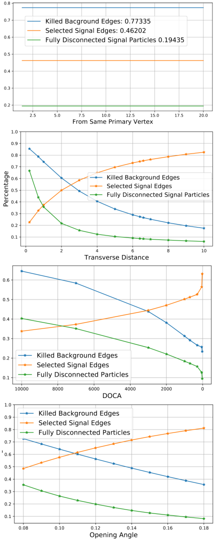


Figure: Fast

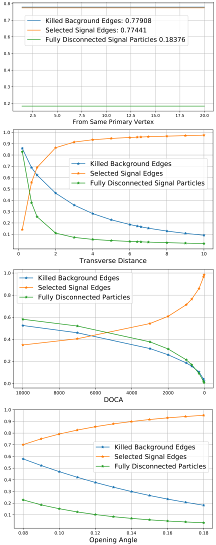


Figure: Full

How many particles are retained after applying the cut on the edges?

A particle is retained if it is connected to at least one other particle through an edge. The signal comprises pairs of particles originating from the same primary vertex of the b hadron ($FromSamePV = 1$), while the remaining particles are considered background.

Simulation	**Saved Signal** (%)	**Lost Background** (%)
Fast inclusive	97.5 ± 0.2	85.8 ± 0.15
Full Inclusive	97.6 ± 0.2	88.1 ± 0.15

Table: Efficiency of the pre selection cut on particles. The signal particles saved by the cut on the edges are the "Saved Signal" while the background particles removed by the cut are the "Lost Background".

Performance Classification: Perfect and Quasi

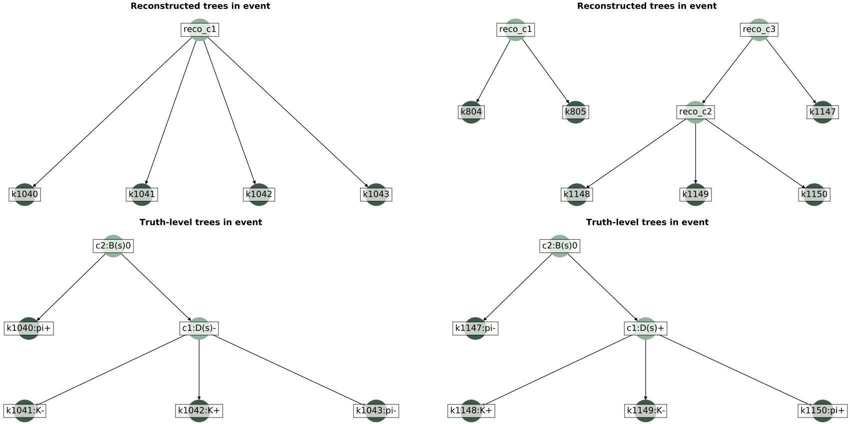
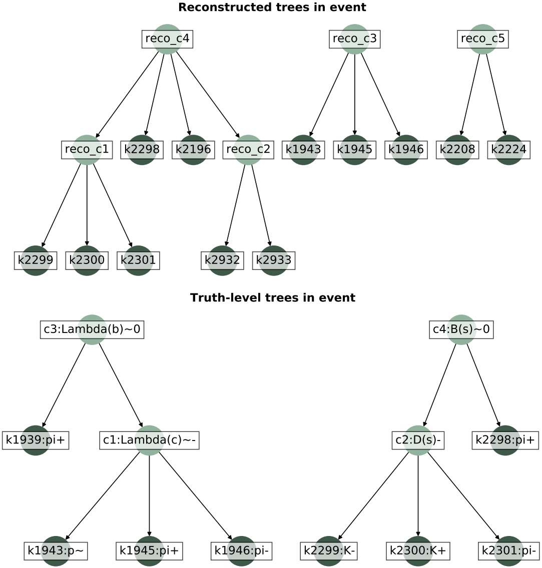
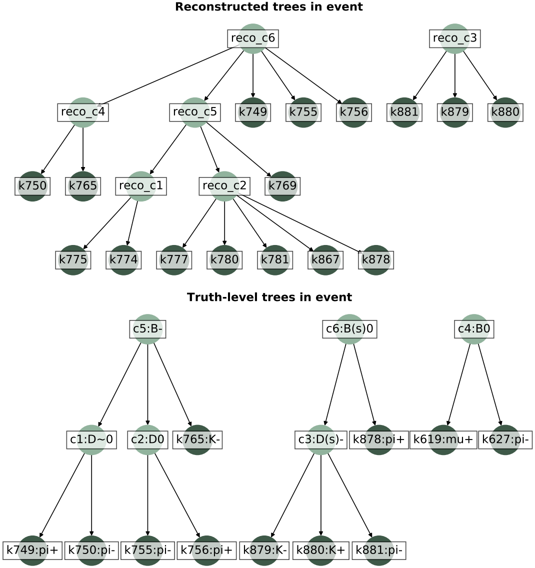


Figure: Quasi (left), Perfect (right)

Performance Classification: Not-Isolated



Performance Classification: Partial

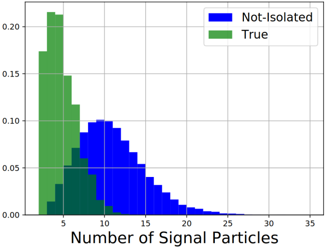


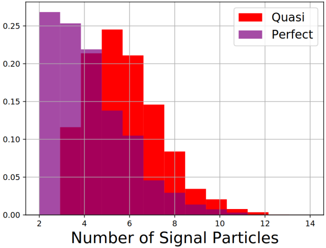
Performance classification of specific decays in the inclusive $b\bar{b}$ sample

	Events	**psr** (%)	**Quasi** (%)	**Not-Isolated** (%)	**Partial** (%)
$B^+ \to \bar{D}^0[K^+\pi^-]\mu^+\nu_\mu$	97	3.6 ± 1.7	5.4 ± 2.1	91.1 ± 3.0	0.0 ± 0.0
$B^+ \to \bar{D}^0[K^+\pi^-]e^+\nu_e$	123	6.0 ± 2.1	3.8 ± 1.6	90.2 ± 2.6	0.0 ± 0.0
$B^0 \to D^-[K^+\pi^-\pi^-]\mu^+\nu_\mu$	31	3.6 ± 3.5	3.6 ± 3.5	92.9 ± 4.9	0.0 ± 0.0
$B^0 \to D^-[K^+\pi^-\pi^-]e^+\nu_e$	40	7.5 ± 4.2	5.0 ± 3.4	87.5 ± 5.2	0.0 ± 0.0
$B^0 \to K^+\pi^-$	14	4.5 ± 4.4	0.0 ± 0.0	95.4 ± 4.4	0.0 ± 0.0

Table: Percentage of perfect, quasi, Not-Isolated and partial signal reconstruction in the fast simulation with 49399 events.

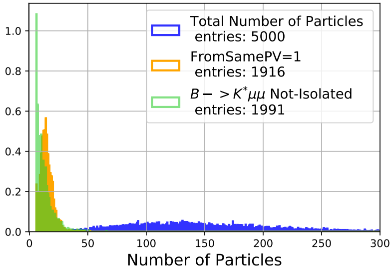
Extra Particles in the inclusive $b\bar{b}$ sample.





How many particles of the same PV DFEI can remove?

In the case of the $B \to K^* \mu\mu$ decay, it has been compared the number of background particles in the not-isolated signals with respect to the total distribution of background particles coming from the same primary vertex ($FromSamePV = 1$) per event. This was done in order to evaluate how many background particles DFEI manages to remove from a PV. Then, these two quantities are further compared with the total number of particles per event.



Training of the LCA module

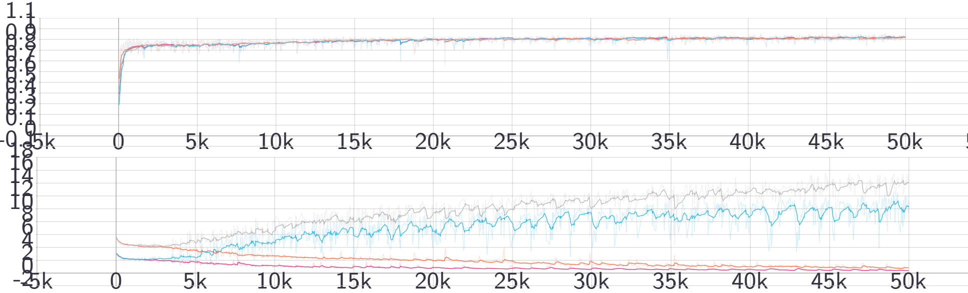


Figure: Total Accuracy (up), loss function (down)