Karlsruhe Institute of Technology

Institute of Experimental Particle Physics

$B_s^0 \to \tau^-\tau^+$ at FCC-ee News

General status and prospects

December 4 2025

Alejandro Q. Triviño

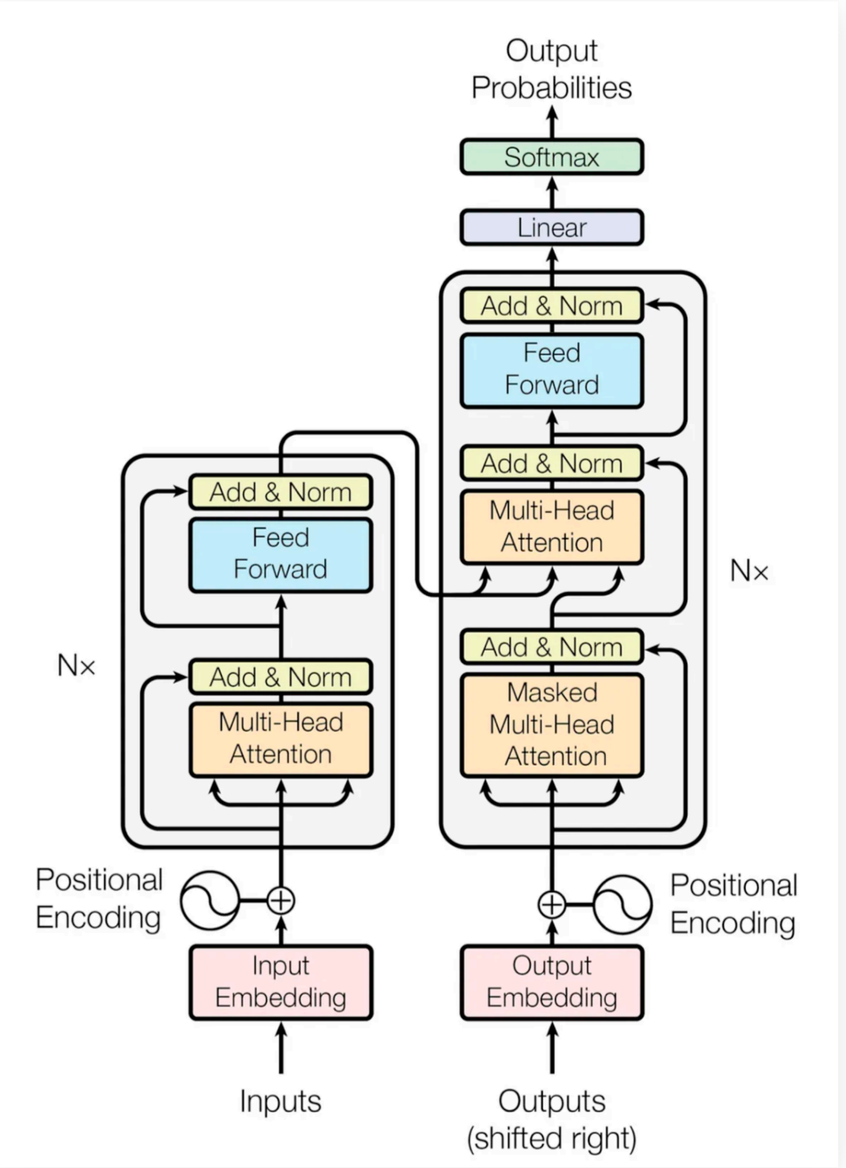
Outline

- Architecture Vanilla Transformer (VT) for classification
- Study of performance of classifier for difference set of input variables
  - Particle information (Taus and Pions)
  - Particle information + Event information
  - Tracks information
  - All information combined = Tracks + Particle + Event
- After selection Particle information + Event information
- Summary

The architecture

Vanilla transformer

- Transformers were developed to make Large Language Models work better, with the ability to “remember” what was important in earlier text.
- The secret of Transformers is the self-attention mechanism (Scaled Dot Product Attention), which finds hidden correlations in the data.
- Our classifier is based on the vanilla Transformer architecture (“*Attention Is All You Need*”)



Monte Carlo Samples

- Official Monte Carlo samples from the **FCC Winter 2023 campaign**
- Fast simulation Pythia8 + EvtGen + Delphes
- **IDEA** detector and **EDM4HEP** data format

Signal, 10M events available

- $B_s \to \tau^+\tau^-, \tau \to 3\pi + \nu_\tau$ (Exclusive)

Background, 500M events for background process

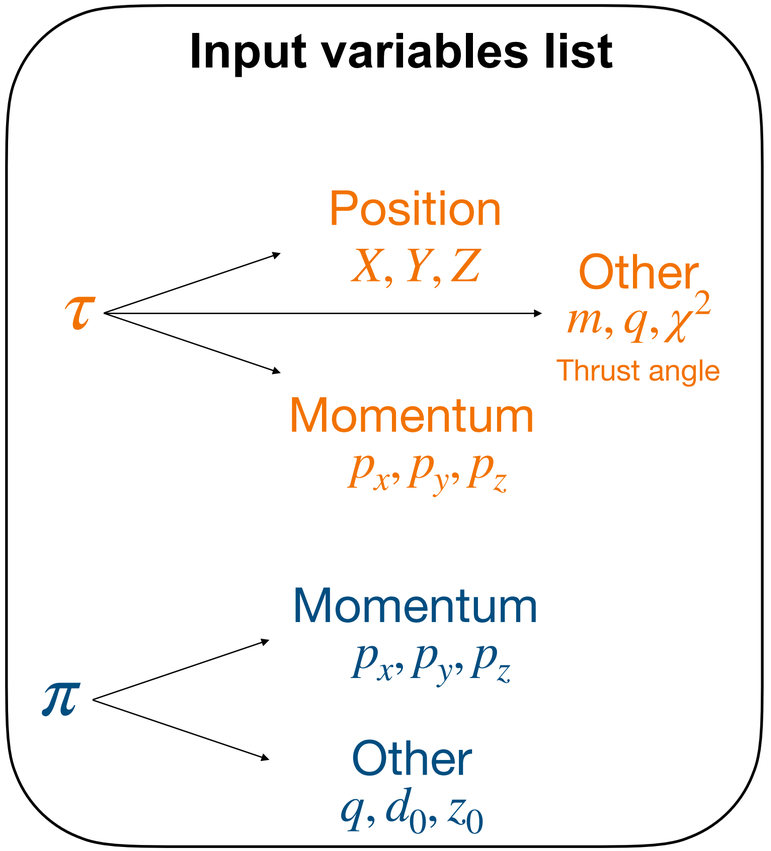
- $Z^0 \to b\bar{b}$ inclusive (Predominant background)
- $Z^0 \to c\bar{c}$ inclusive (Available but not used)
- $Z^0 \to s\bar{s}$ inclusive (Available but not used)
- $Z^0 \to ud$ inclusive (Available but not used)

Particle information (Tau, Pion)

Transformer Based Classifier

Particle information

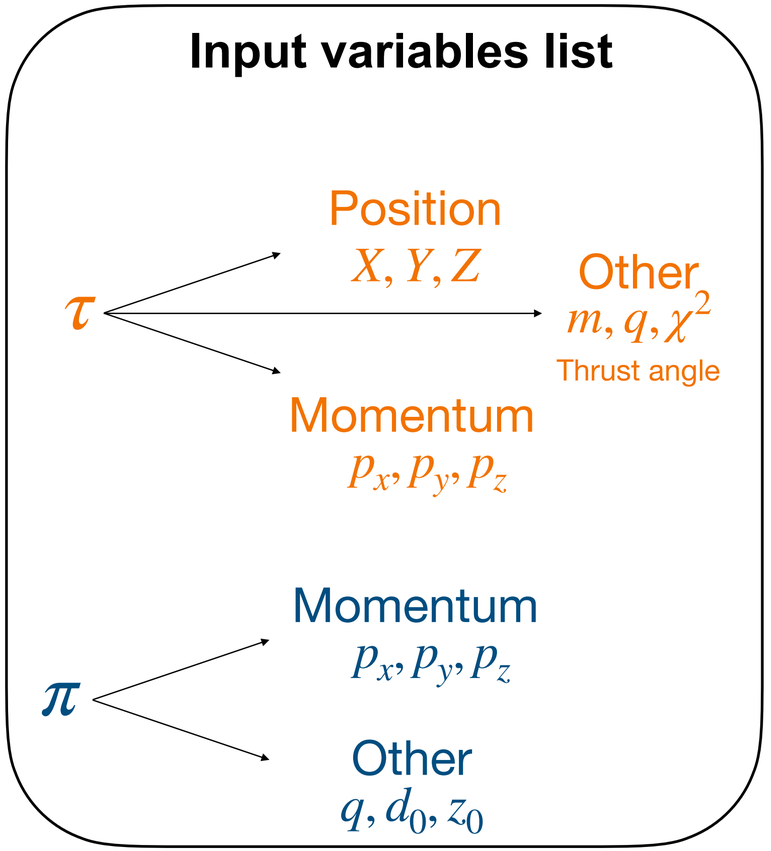
- **Goal:** Build a classifier to distinguish $B_s^0 \to \tau^+\tau^- \to 6\pi + 2\nu_\tau$ events vs $Z^0 \to b\bar{b}$ inclusive background
- **Input features:** Reconstructed information from taus and pions
- Focus on **hadronic taus decays**, which reduce the branching ration to $\mathcal{B}(B_s^0 \to \tau^-\tau^+) \approx 10^{-9}$ (**challenging measurement**)

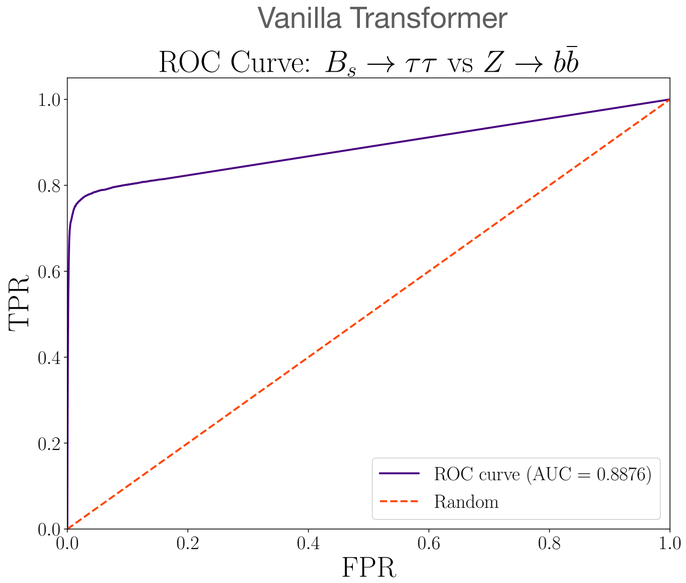


Transformer Based Classifier

Particle information

- **Goal:** Build a classifier to distinguish $B_s^0 \to \tau^+\tau^- \to 6\pi + 2\nu_\tau$ events vs $Z^0 \to b\bar{b}$ inclusive background
- **Input features:** Reconstructed information from taus and pions
- Focus on **hadronic taus decays**, which reduce the branching ration to $\mathcal{B}(B_s^0 \to \tau^-\tau^+) \approx 10^{-9}$ (**challenging measurement**)

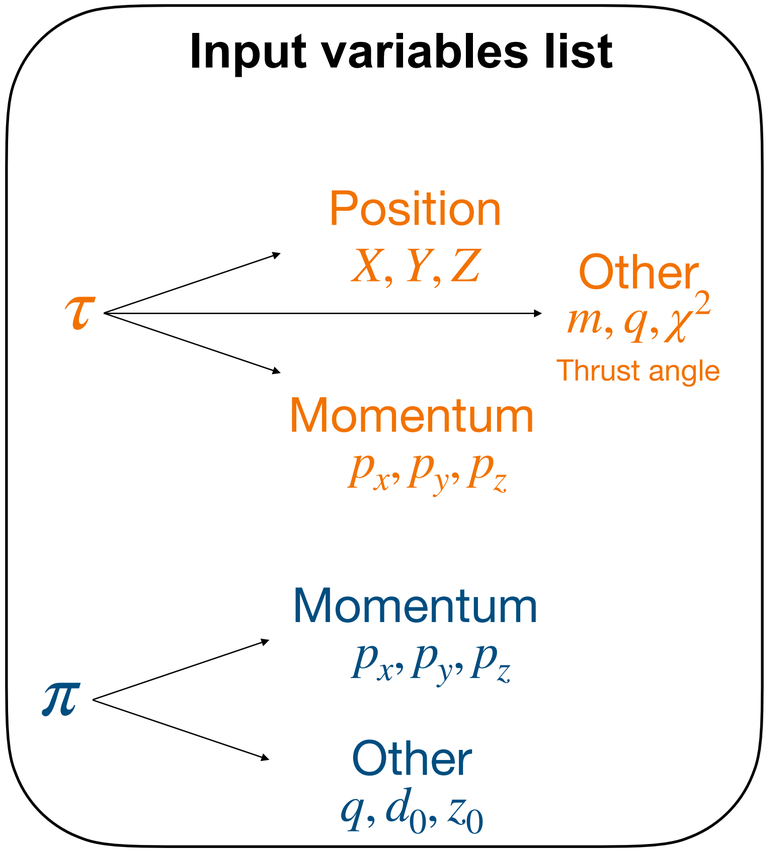


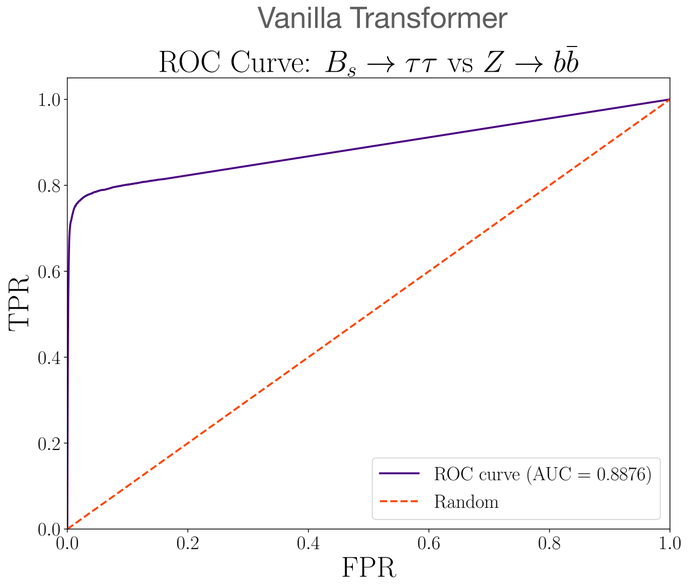


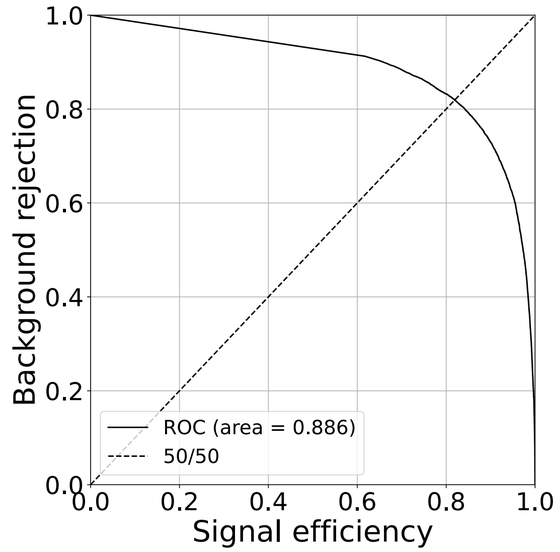
Transformer Based Classifier

Particle information

- **Goal:** Build a classifier to distinguish $B_s^0 \to \tau^+\tau^- \to 6\pi + 2\nu_\tau$ events vs $Z^0 \to b\bar{b}$ inclusive background
- **Input features:** Reconstructed information from taus and pions
- Focus on **hadronic taus decays**, which reduce the branching ration to $\mathcal{B}(B_s^0 \to \tau^-\tau^+) \approx 10^{-9}$ (**challenging measurement**)







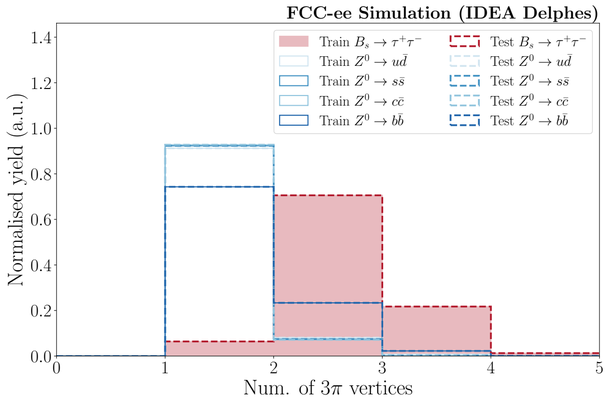
LGATr

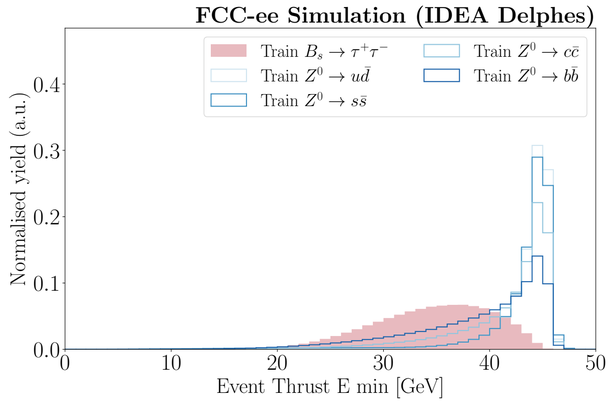
Particle + Event level information

Transformer Based Classifier

Particle information + Event level information

- What happened if we include additional information from the hemispheres?
- Some relevant variables that have shown discrimination power before, like Energy and multiplicity in both hemispheres.





Input variables list

τ
- Position X, Y, Z
- Other m, q, χ^2 Thrust angle
- Momentum p_x, p_y, p_z

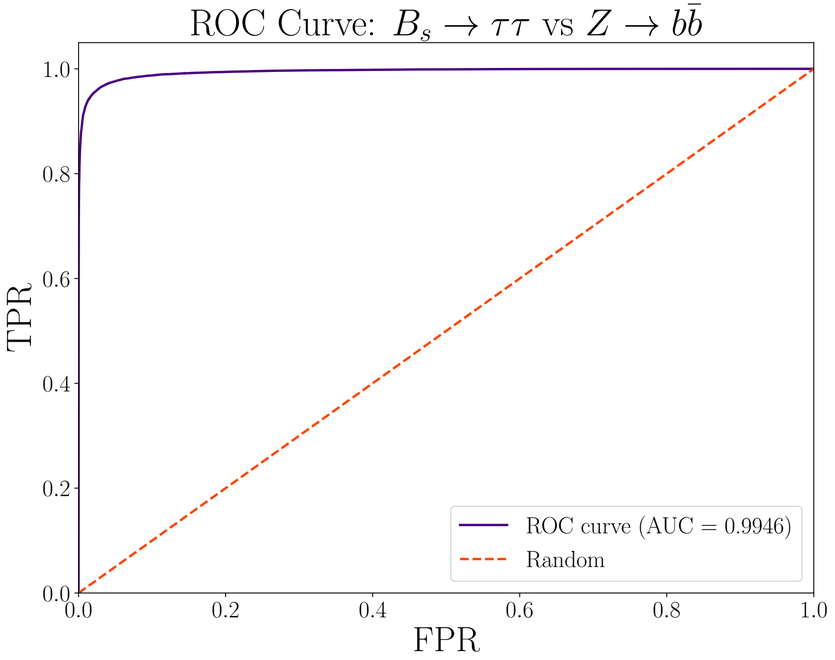
π
- Momentum p_x, p_y, p_z
- Other q, d_0, z_0

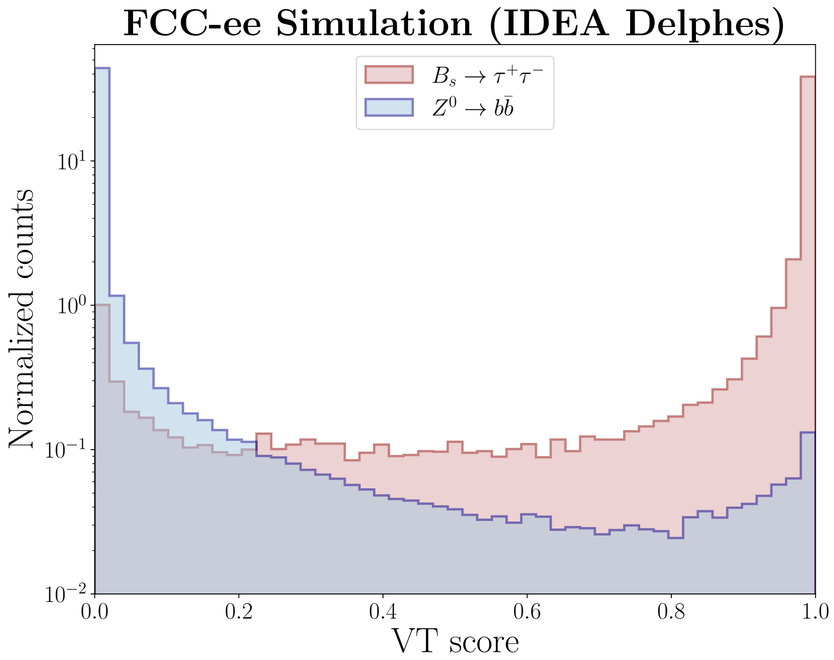
EVT
- Energy $(E_T, E_q, E_N)_{min/max}$
- Multiplicity $(N_q, N_N, N_{\tau\to 3\pi})_{min/max}$

Transformer Based Classifier

Particle information + Event level information

- Trained in 500K signal events and 500K background events
- Classification improve notably after including event level information, very useful information since the event topology of the process
- Similar performance as best BDT result, but still window of improvement



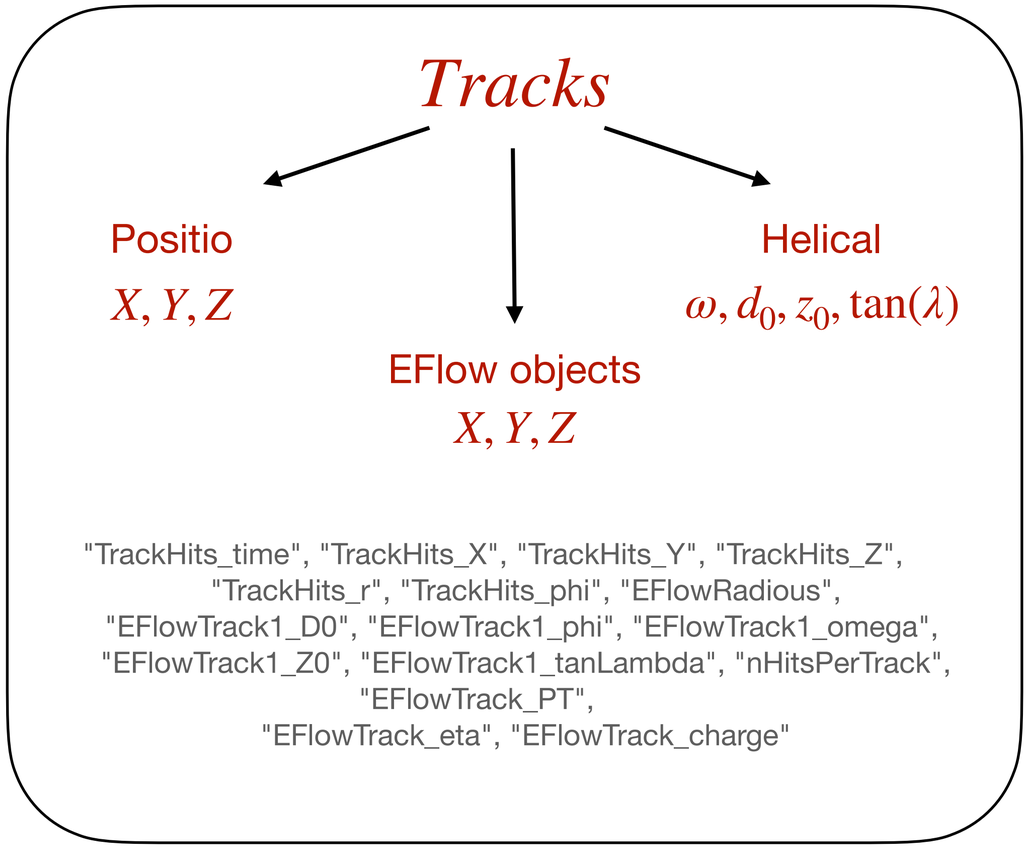


Tracks information

Transformer Based Classifier

Tracks information

Input variables list



Transformer Based Classifier

Tracks information

Input variables list

Tracks

Positio

X, Y, Z

Helical

$\omega, d_0, z_0, \tan(\lambda)$

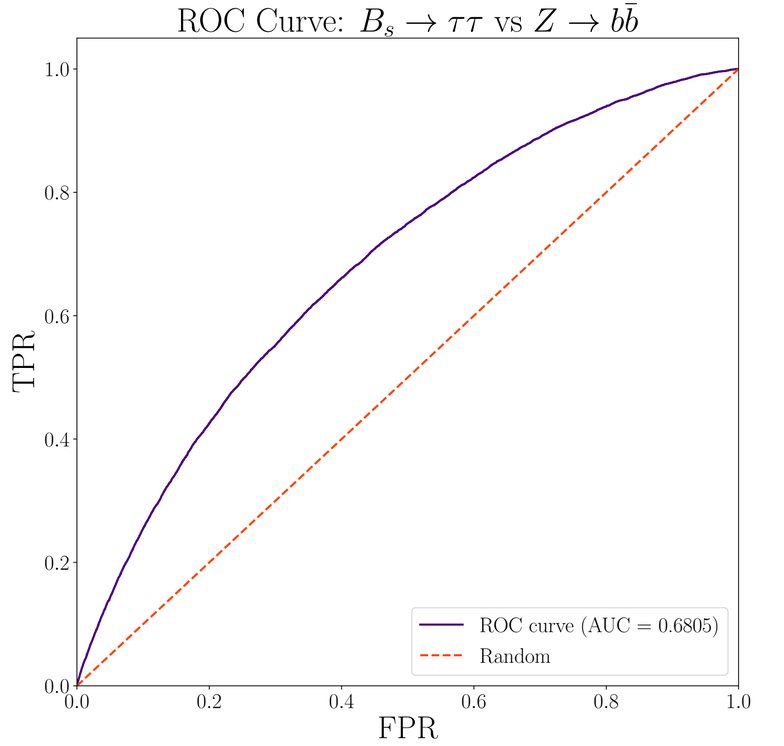
EFlow objects

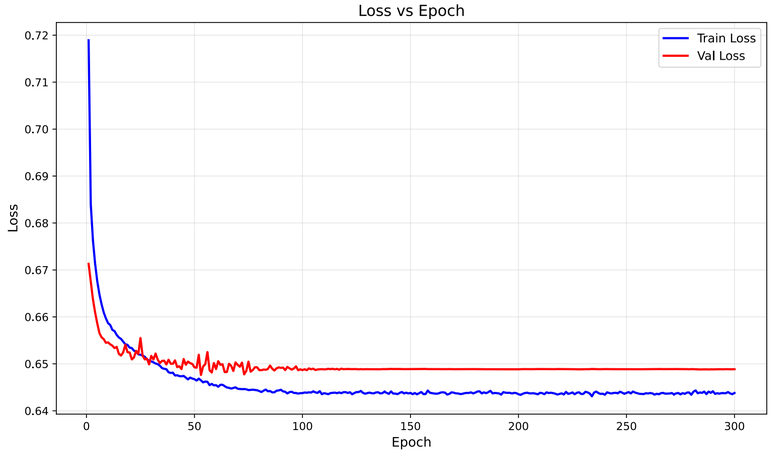
X, Y, Z

"TrackHits_time", "TrackHits_X", "TrackHits_Y", "TrackHits_Z",
"TrackHits_r", "TrackHits_phi", "EFlowRadious", "EFlowState_begin",
"EFlowState_end", "EFlowdxQ_begin", "EFlowdxQ_end",
"EFlowTracker_begin",
"EFlowTracker_end", "EFlowTrack1_D0", "EFlowTrack1_phi",
"EFlowTrack1_omega",
"EFlowTrack1_Z0", "EFlowTrack1_tanLambda", "nHitsPerTrack",

1M signal events

1M background events





All categories

Transformer Based Classifier

All information

Input variables list

All categories

τ π

Particles

EVT

Event level

$Tracks$

Tracks info

Transformer Based Classifier

All information

Input variables list

All categories

τ π

Particles

EVT

Event level

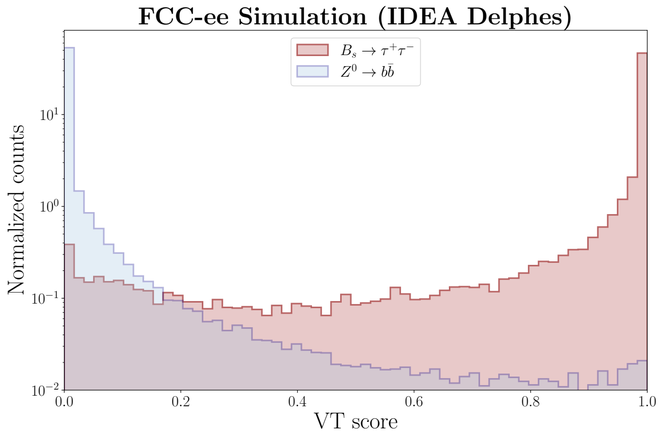
Tracks

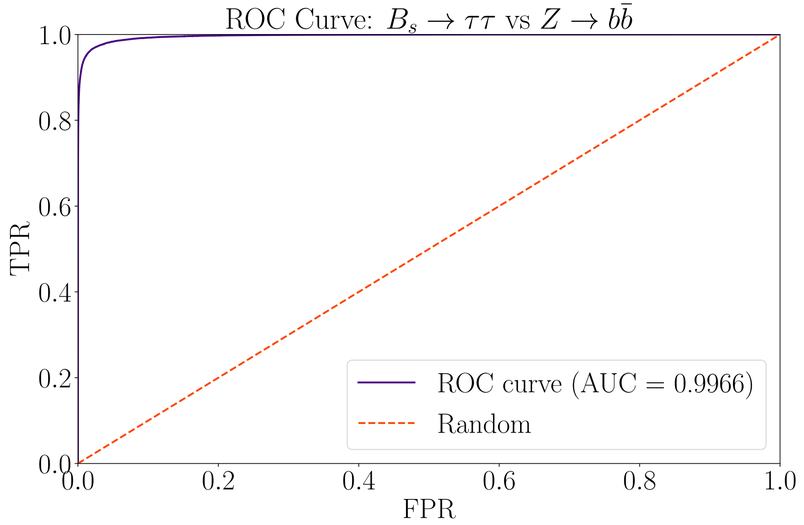
Tracks info

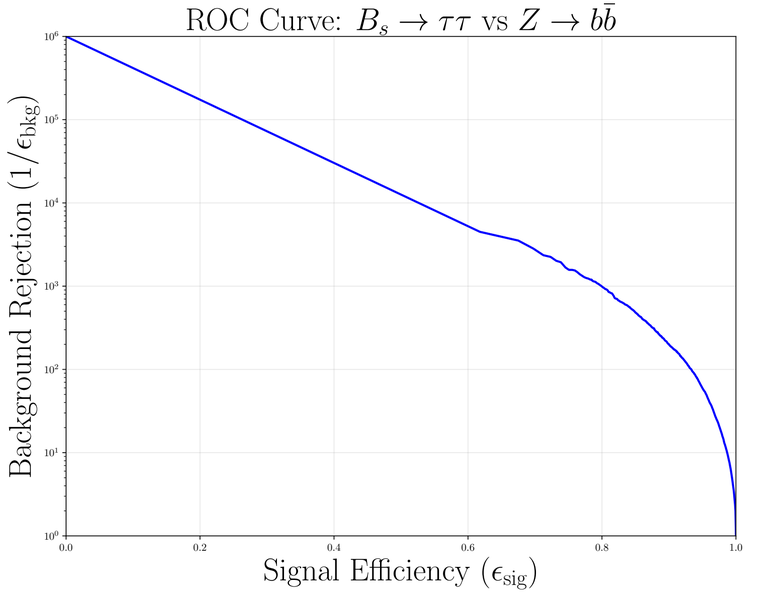
$10^3 \rightarrow 10^4$ Background rejection achieve

1M signal events

1M background events





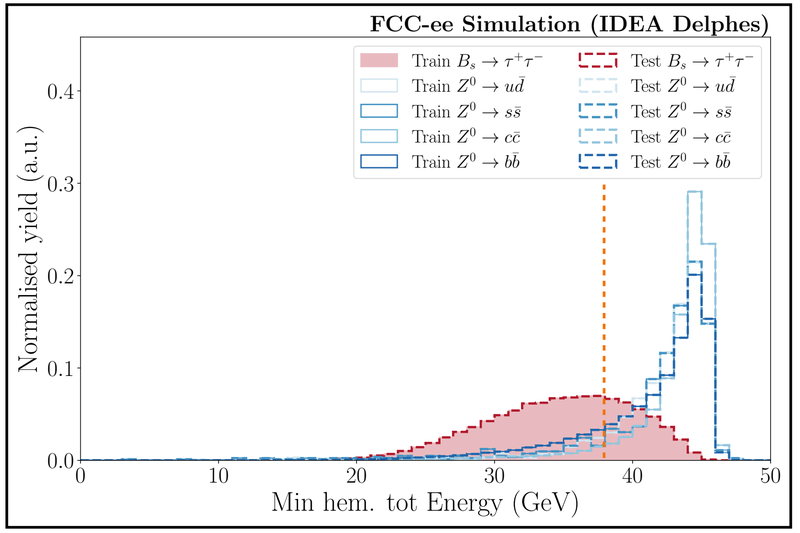


Reduce phase space

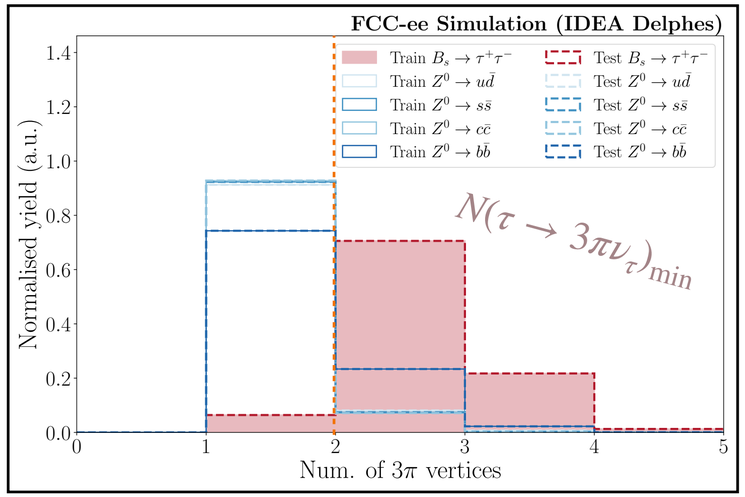
Event selection

Sensitive variables

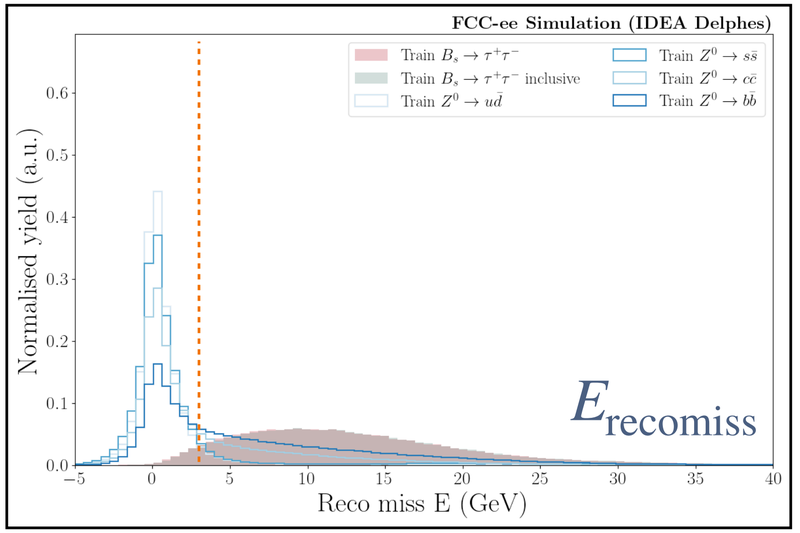
- Among all variables studied in this analysis, **five key variables** show the highest discriminating power, significantly reducing the background efficiency by several orders of magnitude after applying basic selection cuts.



Total energy in the signal hemisphere



Number of three pions reconstructed vertices (from the tau decay)

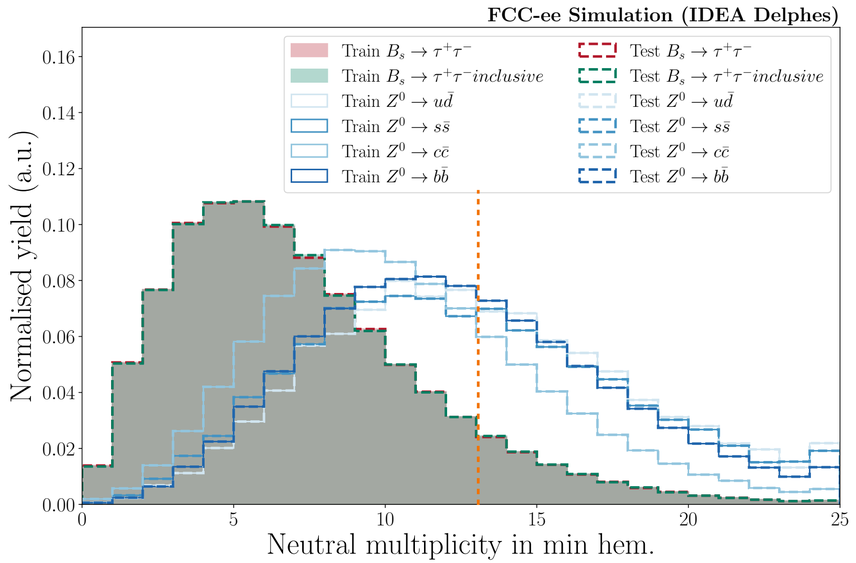


Miss reconstructed energy in the signal hemisphere

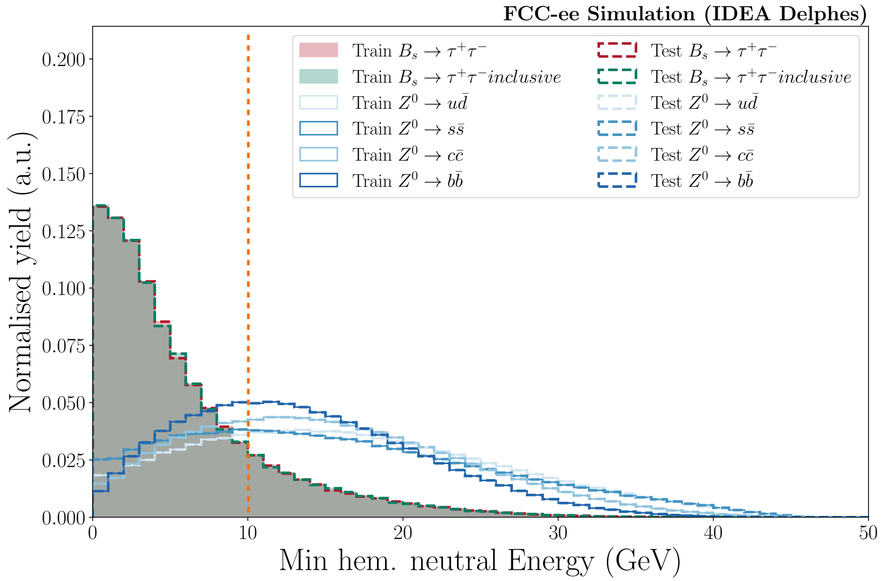
Event selection

Sensitive variables

- Among all variables studied in this analysis, **five key variables** show the highest discriminating power, significantly reducing the background efficiency by several orders of magnitude after applying basic selection cuts.



Number of neutral particles in the signal hemisphere



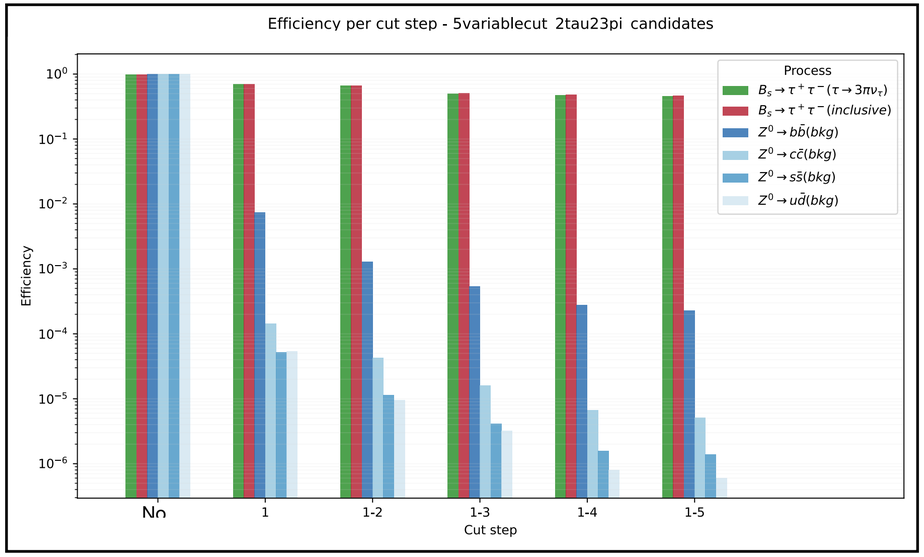
Total reconstructed energy of neutral particles

Event selection

Classical selection

- Apply cuts on the most **sensitive variables** to the existing samples
- Reduce background efficiency three order of magnitudes, while maintaining about 50% signal efficiency
- $Z^0 \to b\bar{b}$ inclusive is the most challenging background to separated in this analysis
- A Machine Learning implementation required to improve background rejection

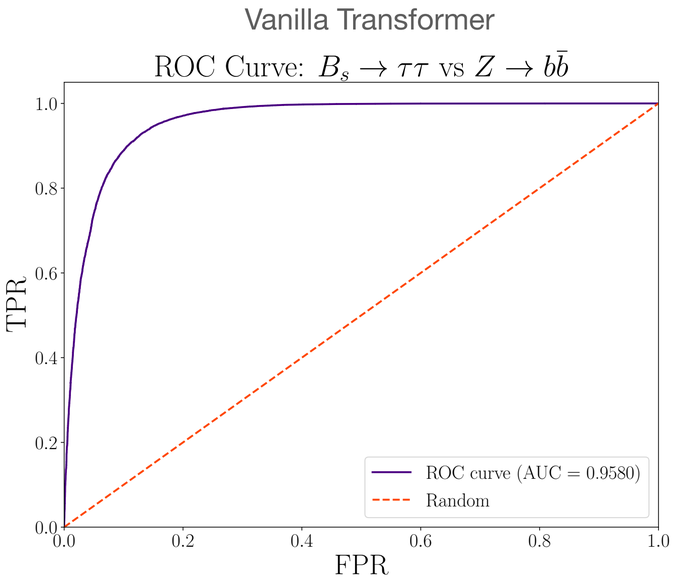
Cut	Variable
1	$N(\tau \to 3\pi\nu_\tau)_{\text{min}} = 2$
2	$E_{\text{recomiss}} > 3.5\,\text{GeV}$
3	$E_{\text{min}} < 38\,\text{GeV}$
4	$E_{\text{neutral}} < 10\,\text{GeV}$
5	$N_{\text{neutral}} < 13$

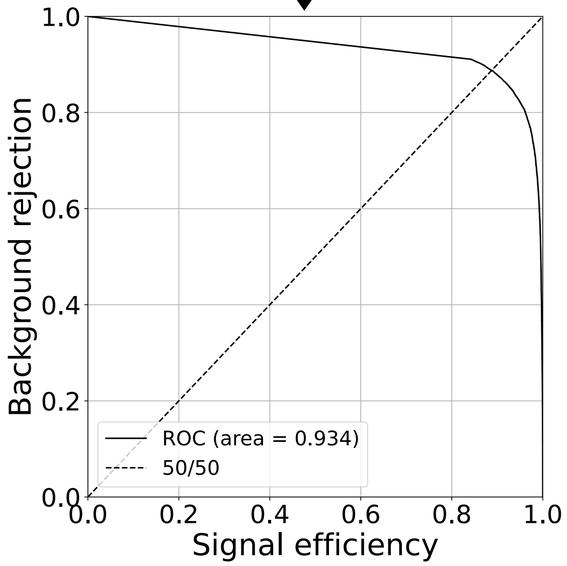


Transformer Based Classifier

Constraint phase space

- Implement vanilla transformer in new phase space
- Increase performance from previous LGTAr result, background efficient
- After classical selection and classification we achieve a background rejection:





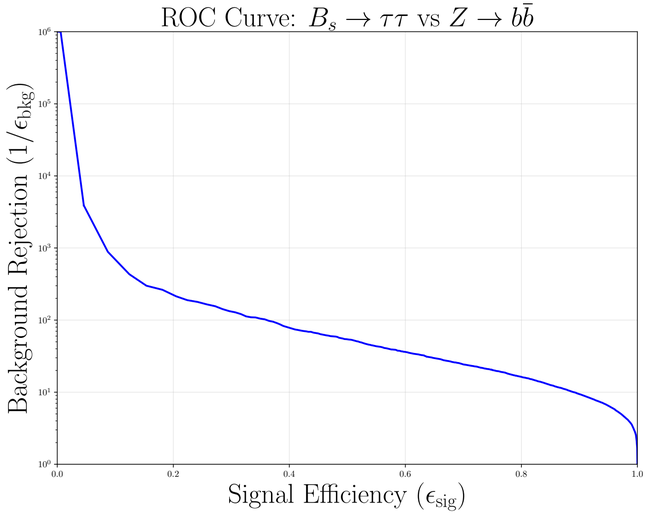
LGATr

Transformer Based Classifier

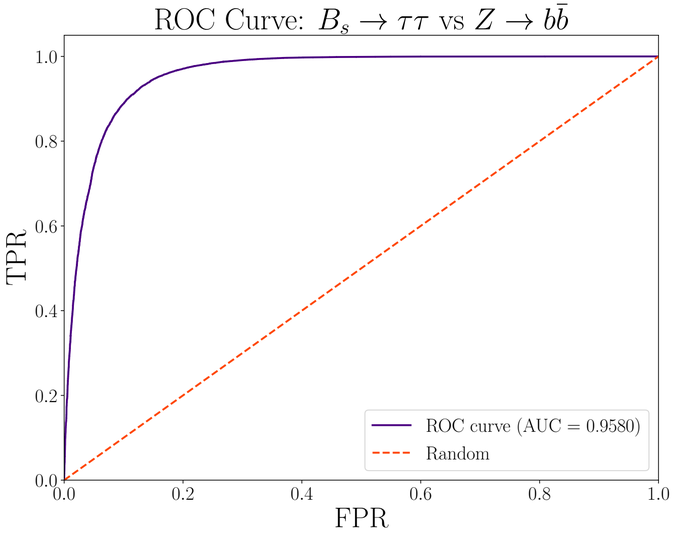
Constraint phase space

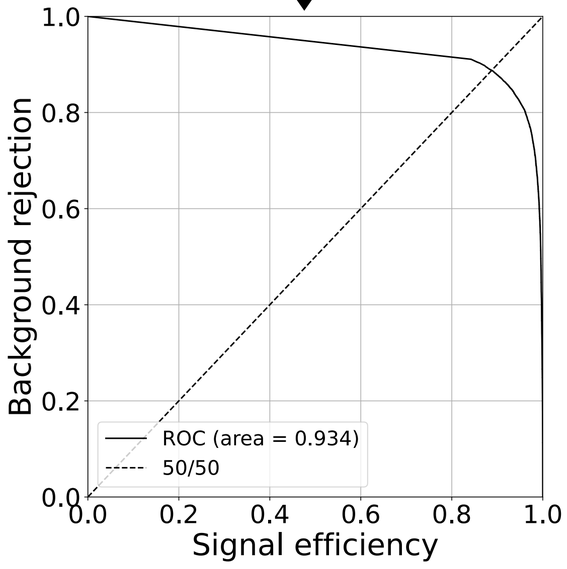
- Implement vanilla transformer in new phase space
- Increase performance from previous LGTAr result, background efficient
- After classical selection and classification we achieve a background rejection:

$10^5 \rightarrow 10^6$ Background rejection achieve



Vanilla Transformer





LGATr

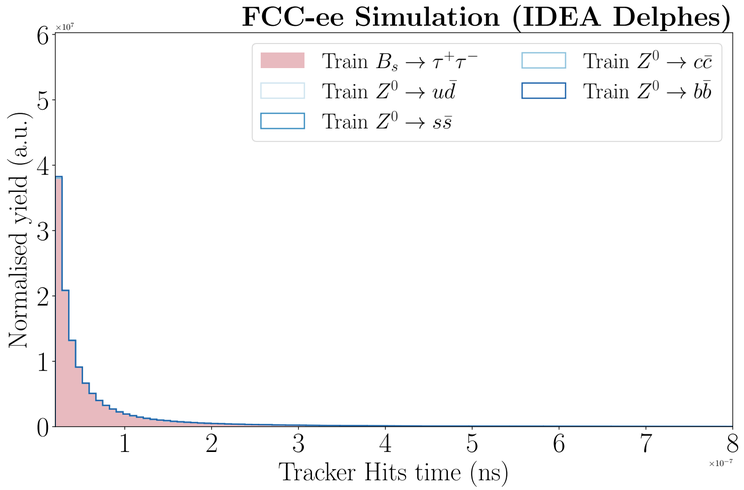
Transformer Based Classifier

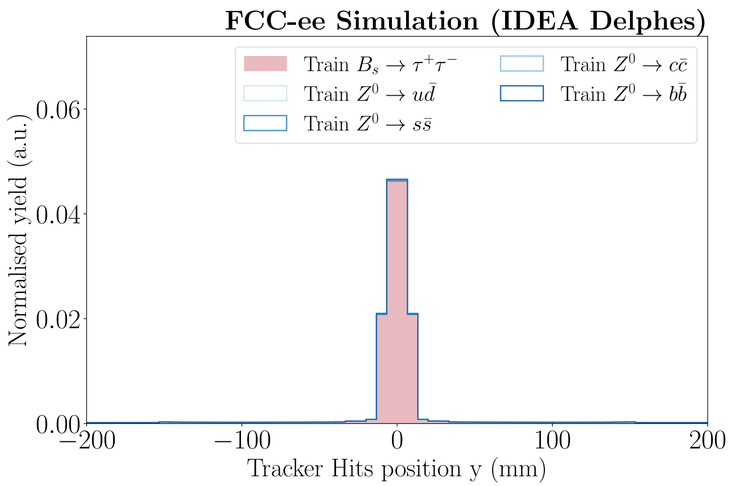
Summary

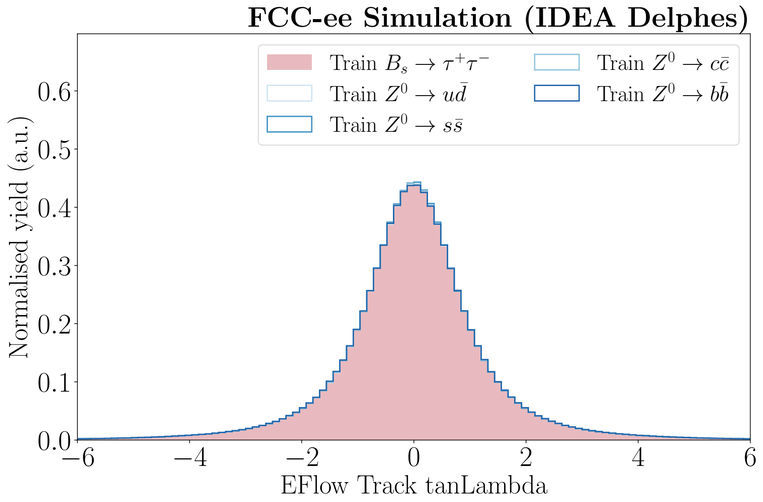
- Vanilla transformer implementation
- First approach **without classical** selection
  - Signal efficiency $\approx 80\,\%$ → Background rejection of 10^3
  - Signal efficiency $\approx 50\,\%$ → Background rejection of 10^4
- Second approach **with classical** selection
  - Signal efficiency $\approx 45\,\%$ → Background rejection of 10^5
  - Signal efficiency $\approx 30\,\%$ → Background rejection of 10^6

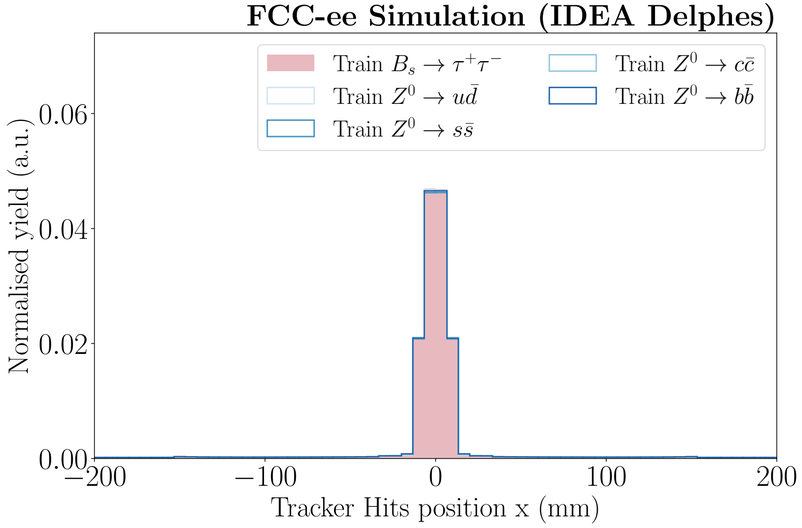
Backup

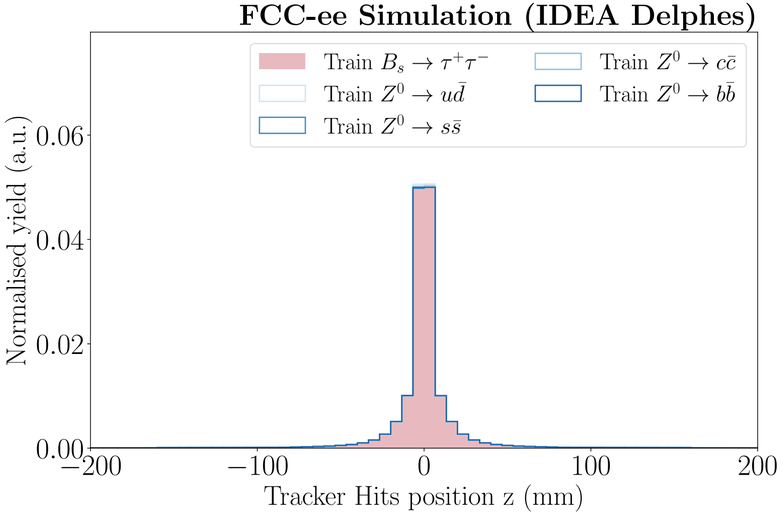
Input variables Tracks

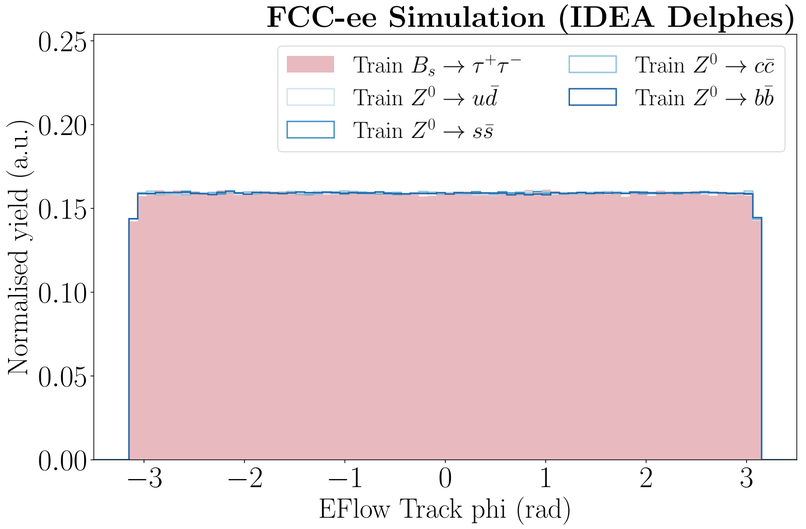




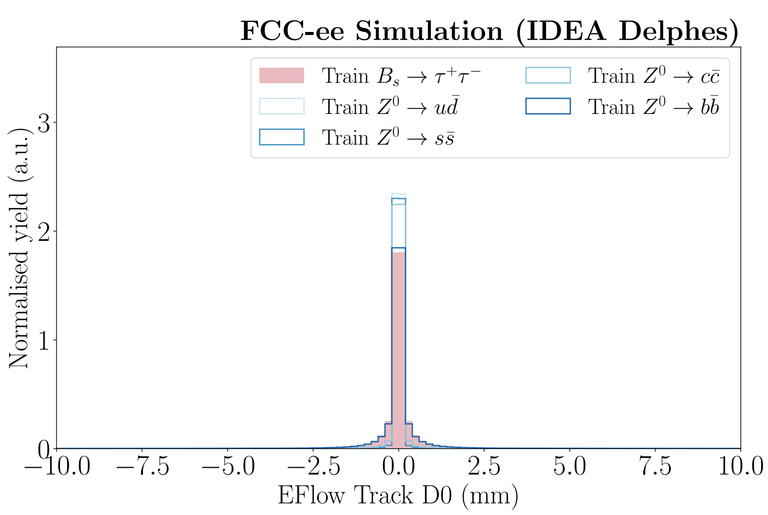


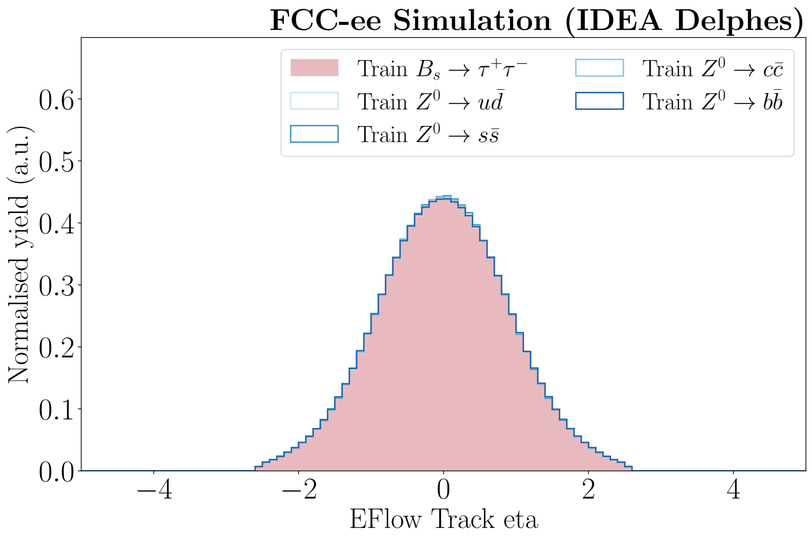


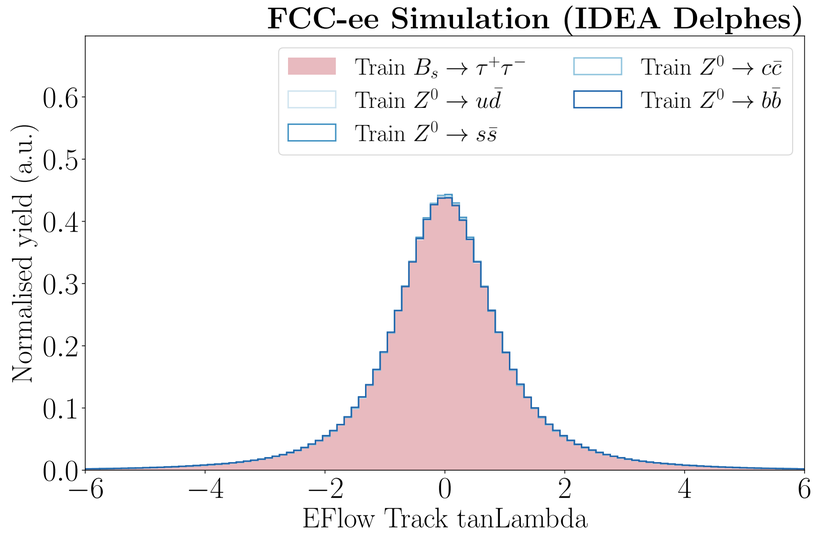


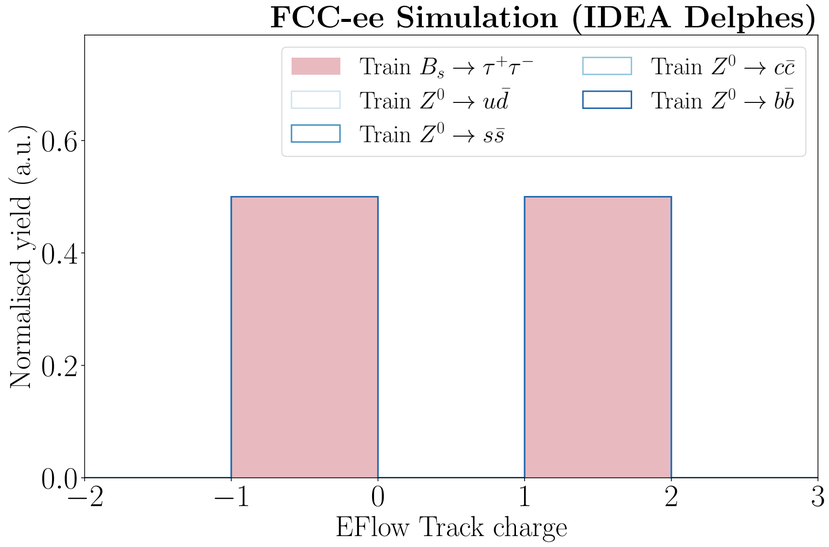


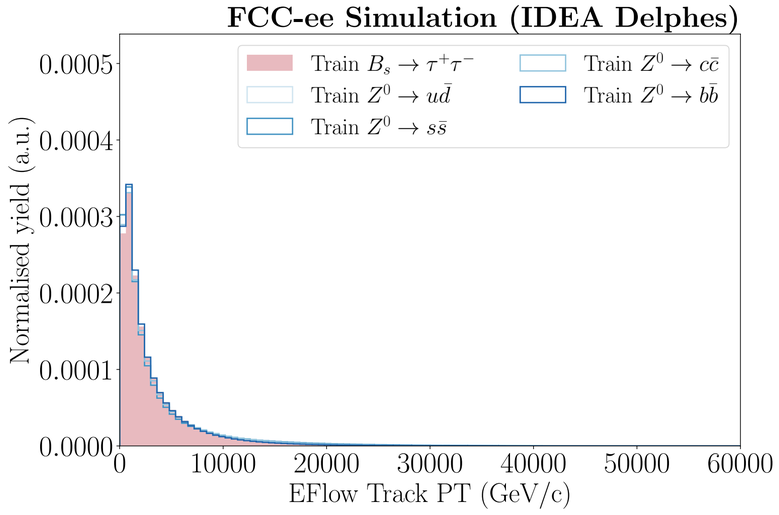
Input variables Tracks

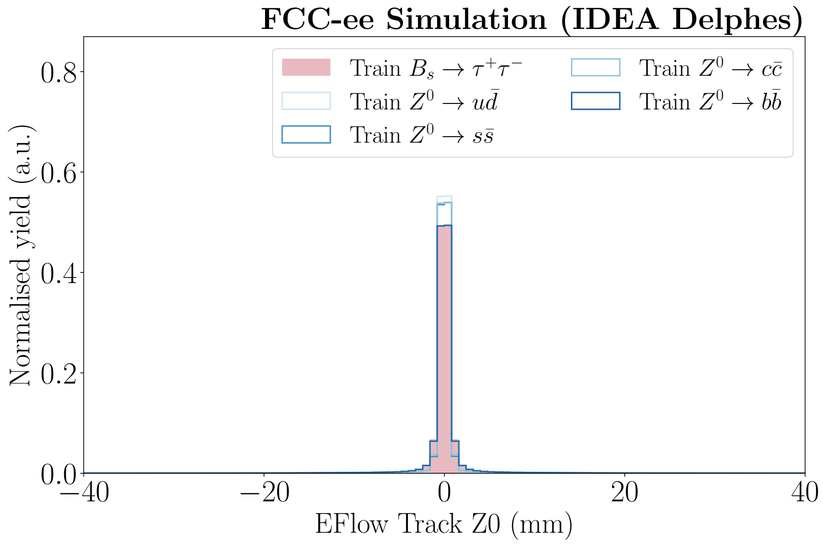












Input variables Tracks