Optimizing Signal-to-Background for $e \to \tau$ Transitions via Leptoquarks at the Electron-Ion Collider

Bardh Quni
Supervisor: Dr. Wouter Deconinck

Faculty of Science
Department of Physics and Astronomy

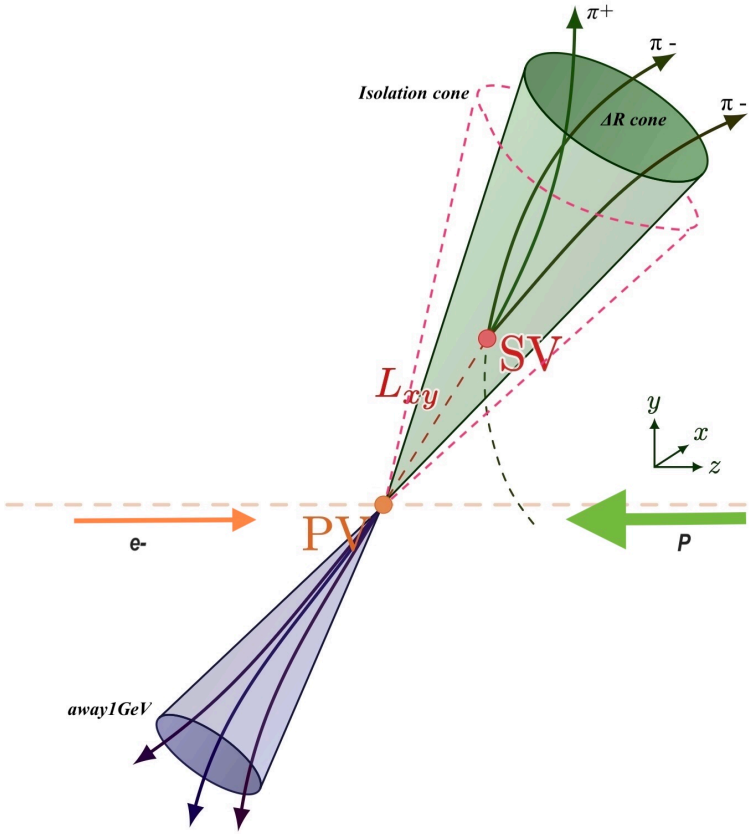
February 10, 2026

University of Manitoba

Contents

Introduction

- Analysis carried for CLFV in DIS-like events mediated by LQs : e + p $\rightarrow \tau$ + X using the highest available center-of-mass energy configuration of $\sqrt{s} = 140$ GeV, corresponding to 18 GeV electron and 275 GeV proton beams.
- Current focus into 3-prong τ decays : $\tau \rightarrow \pi^-\pi^+\pi^-\nu_\tau$



Introduction

Track Parameters:

- CentralCKFTrackParameters.loc.a
- CentralCKFTrackParameters.loc.b
- CentralCKFTrackParameters.θ
- CentralCKFTrackParameters.ϕ
- CentralCKFTrackParameters.qOverP

Particle Information:

- ReconstructedChargedParticles.charge
- ReconstructedChargedParticles.PDG

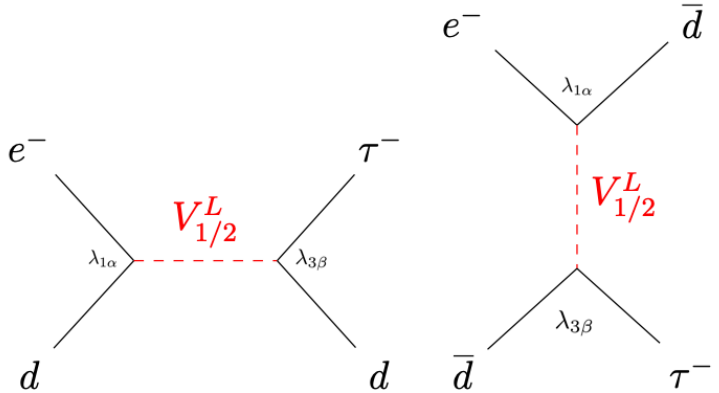
Vertex Information:

- CentralTrackVertices.position.x
- CentralTrackVertices.position.y
- CentralTrackVertices.position.z
- PrimaryVertices_objIdx.index

LQ5=$V_{1/2}^{L}$ Signal and DIS Background

- Presented results regarding the signal are based on s(left) and u(right) channel contributions for $e^- + q_i \rightarrow \mathrm{V}_{1/2}^{L} \rightarrow \tau^- + q_j$ mediated by the Leptoquark(LQ) $V_{1/2}^{L}$

$$\sigma_{|F|=2} = \frac{s}{32\pi}\left[\frac{\lambda_{1\alpha}\lambda_{3\beta}}{M_{\mathrm{LQ}}^2}\right]^2 \int_0^1 dx \int_0^1 dy \left\{x q_\alpha(x,Q^2) f(y) + x\bar{q}_\beta(x,Q^2) g(y)\right\} \tag{1}$$



There are more leptoquarks to look into ...

- S_0^L, S_0^R, $\tilde{S}_0^R$, S_1^L, $S_{1/2}^L$, $S_{1/2}^R$, $\tilde{S}_{1/2}^L$
- V_0^L, V_0^R, $\tilde{V}_0^R$, V_1^L, $V_{1/2}^R$, $\tilde{V}_{1/2}^L$

Consecutive cuts on $V_{1/2}^{L}$, NC, CC, SIDIS events

The selection criteria to identify $e \to \tau$ events includes:

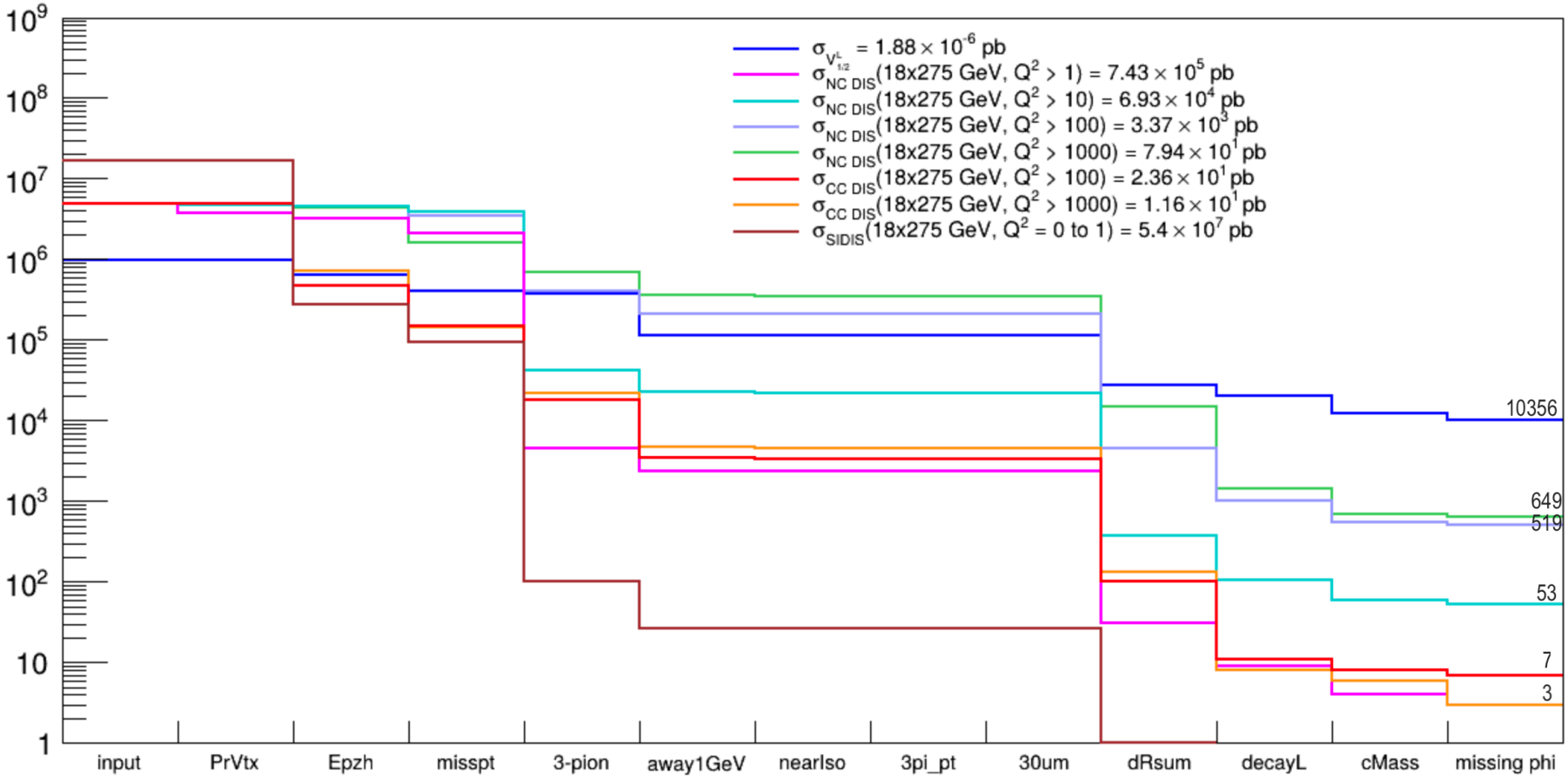
- **PrVtx**: Primary vertex;
- **Epzh**: $\sum_h (E - p_z) > 18$ GeV, where E and p_z are the energy and the z-component of the 3-momentum of the final state particles, and the summation is over all detected hadrons;
- **missing** p_T : $1 < p_T^{miss} < 9$ GeV, where the lower limit is to suppress events with small missing p_T, and the upper limit is to suppress NC and CC events with large missing p_T ;
- **3-pion** : 3 charged pions in a $\Delta R < 1.0$ cone, where R is cone radius in $(\phi - \eta)$ space, $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}$;
- **away1GeV**: p_T sum of all tracks on the away-side of the 3π candidate, $\sum_{\Delta\phi(-p_{3\pi})<1.0} p_T$ is $>$ 1GeV;
- **nearIso**: p_T sum in a cone around the 3π candidate, $\sum_{\Delta R(p_{3\pi})<1.0} p_T$ is $<$ 3.0 GeV;
- **3pi_pt**: p_T sum of the 3 charged-pion, $p_T(3\pi)$, is $>$ 3.0 GeV;
- **30μm**: candidate decay length reconstructed from 3 charged pions is $>$ 30μm ;

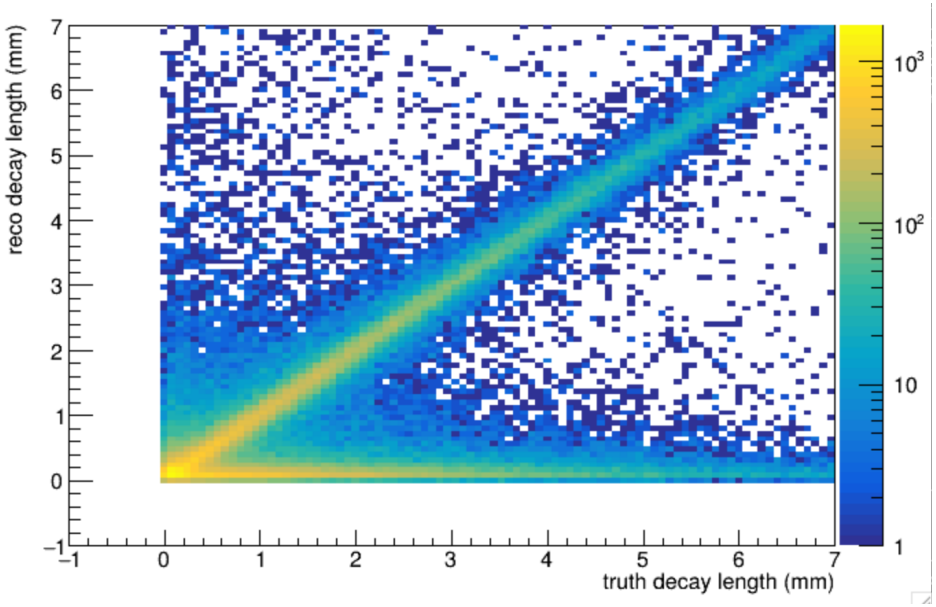
Consecutive cuts on $V^L_{1/2}$, NC, CC, SIDIS events

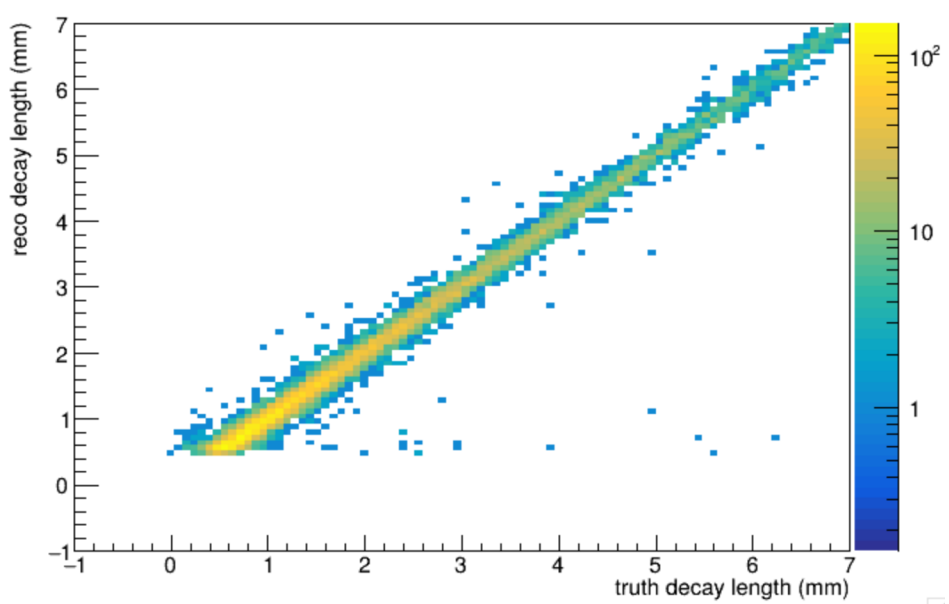
- **dRsum**: sum of the "distances" in $(\phi - \eta)$ space of the 3 charged pions decay vectors $\Delta R = \Delta R_{1,2} + \Delta R_{1,3} + \Delta R_{2,3}$ is $<$ 0.4. Decay vector points from the primary vertex to secondary vertex;
- **decayL**: average of the reconstructed decay length from 3π candidate, $dl = (dl_{1,2} + dl_{1,3} + dl_{2,3})/3$, is $>$ 0.5 mm;
- **cMass**: $\sqrt{M^2_{3\pi} + p^2_{3\pi} sin^2\theta} + p_{3\pi} sin^2\theta <$ 1.8 GeV, $M_{3\pi}$ is mass rconstructed from 3π, while θ is the angle between the reconstructed decay direction and the 3π momentum direction;
- **missing phi**: $\Delta\phi$ between $p_{3\pi}$ and p_T^{miss} is $<$ 1.0.

Consecutive cuts on $V_{1/2}^{L}$, NC, CC, SIDIS events

- Using $V_{1/2}^{L}$ as the signal and the DIS background event samples (see Figure below) from *root://dtn-eic.jlab.org//volatile/eic/EPIC/RECO/25.10.0*, we see that only NC DIS and CC DIS events meet all the selection criteria.



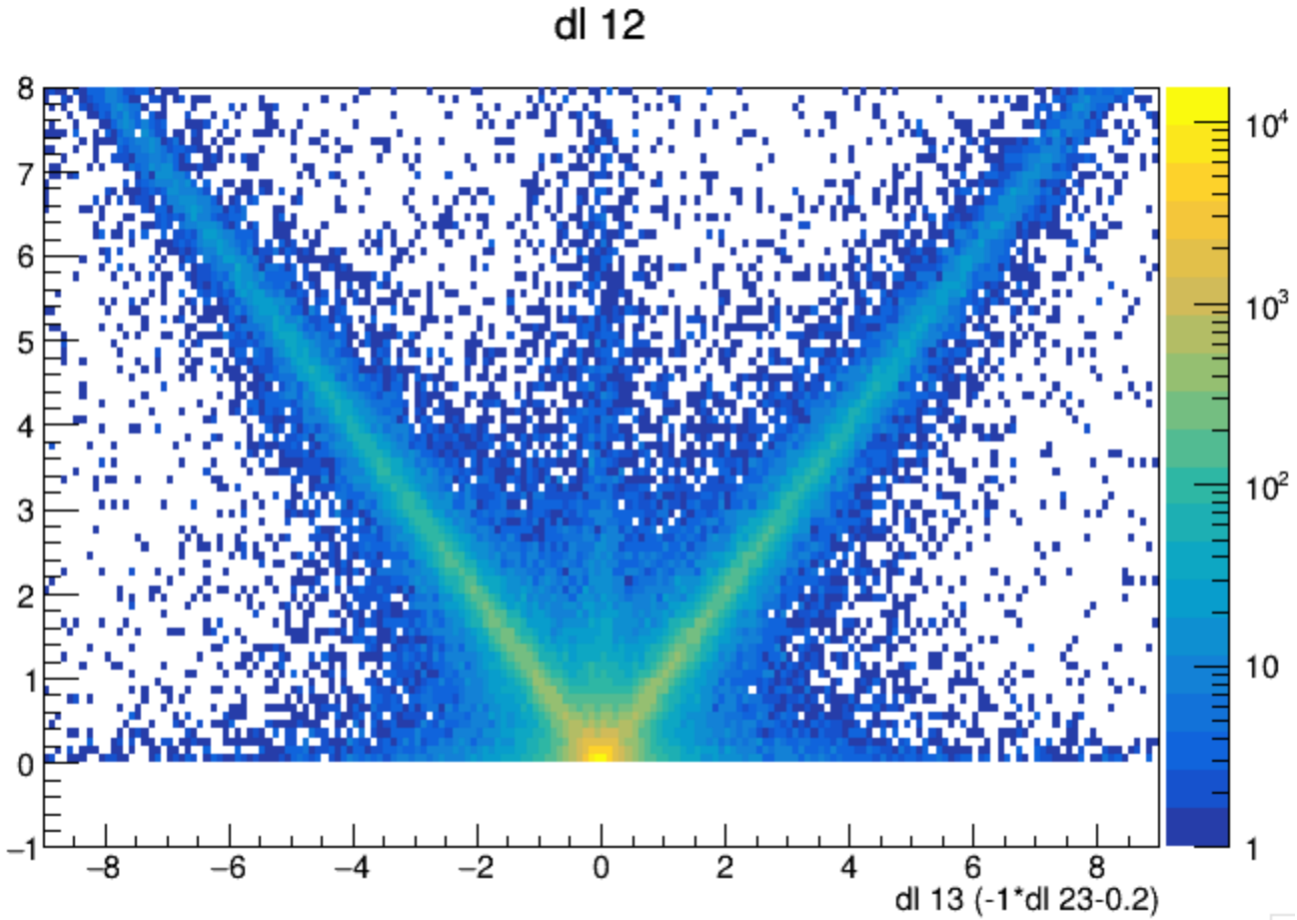




- The final reconstructed decay length (right), after the last cut has been applied, is presented as a comparison between the true (MC) and reconstructed (RC) decay lengths at the ePIC detector level.

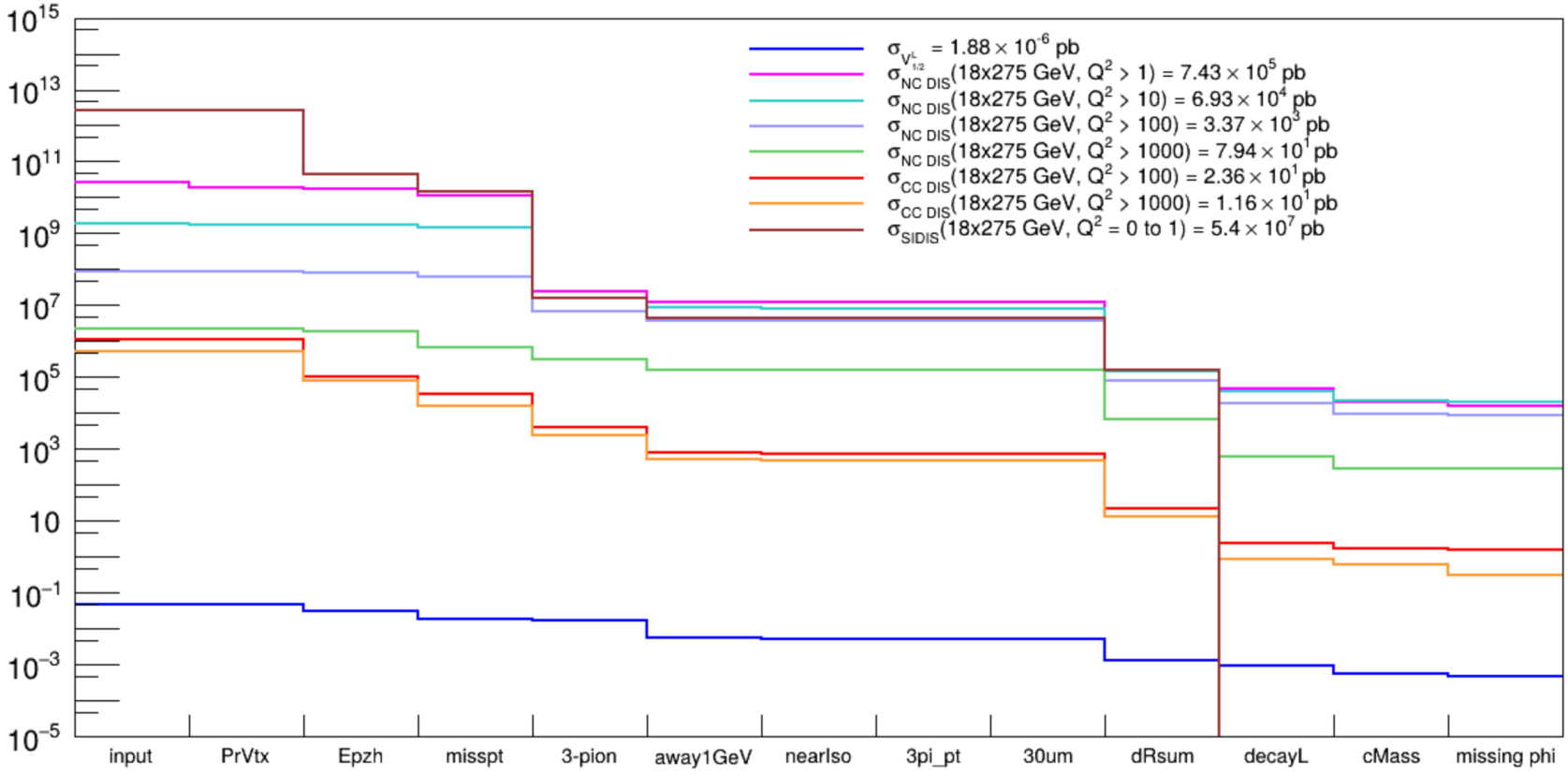
Vertex Identification

- From the simulation of the 3-prong decay mode of the τ lepton, we reconstruct three "intermediate" vertices.
- Coincidence between two out of three vertices (left or right) is usually enough to indicate a "3-prong" secondary vertex.



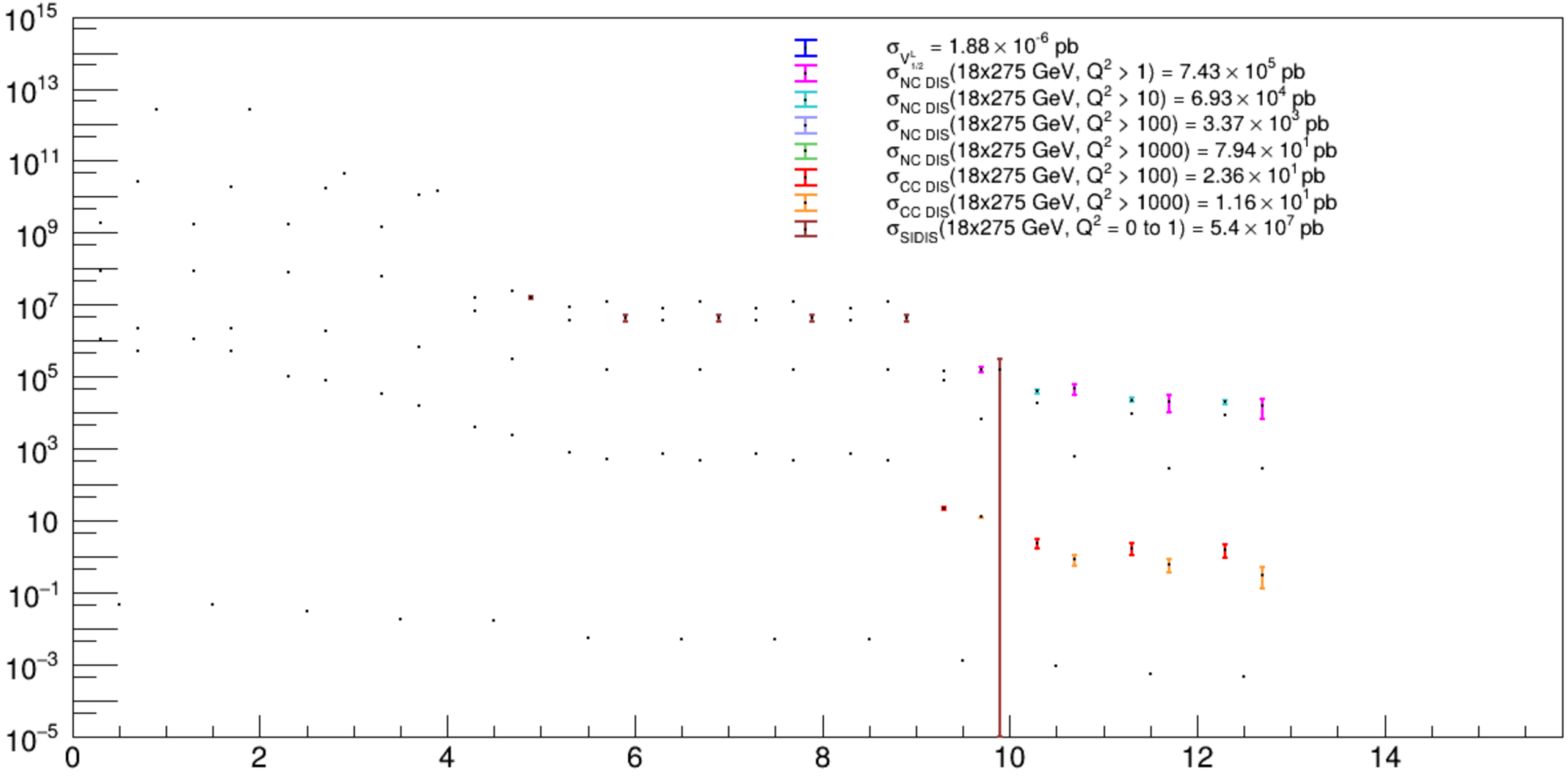
Scaling for LQ5=$V^L_{1/2}$, NCDIS, CCDIS, and SIDIS

- Number of events that survived each selection cut, for an integrated luminosity of $\mathcal{L} = 100\ \mathrm{fb}^{-1}$.



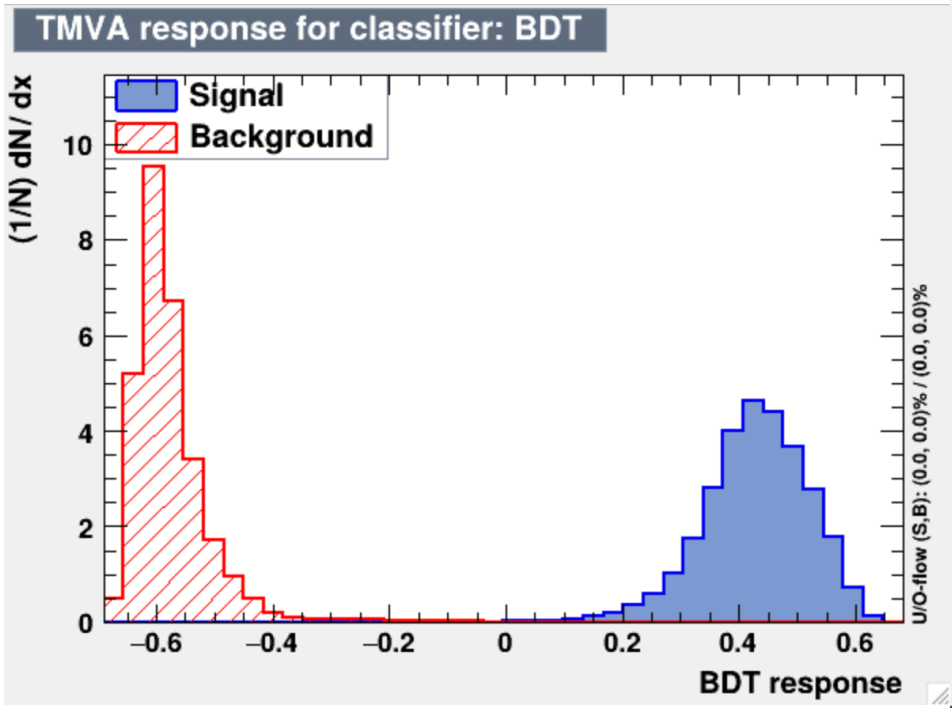
Error bars for LQ5=$V_{1/2}^{L}$, NCDIS, CCDIS, and SIDIS

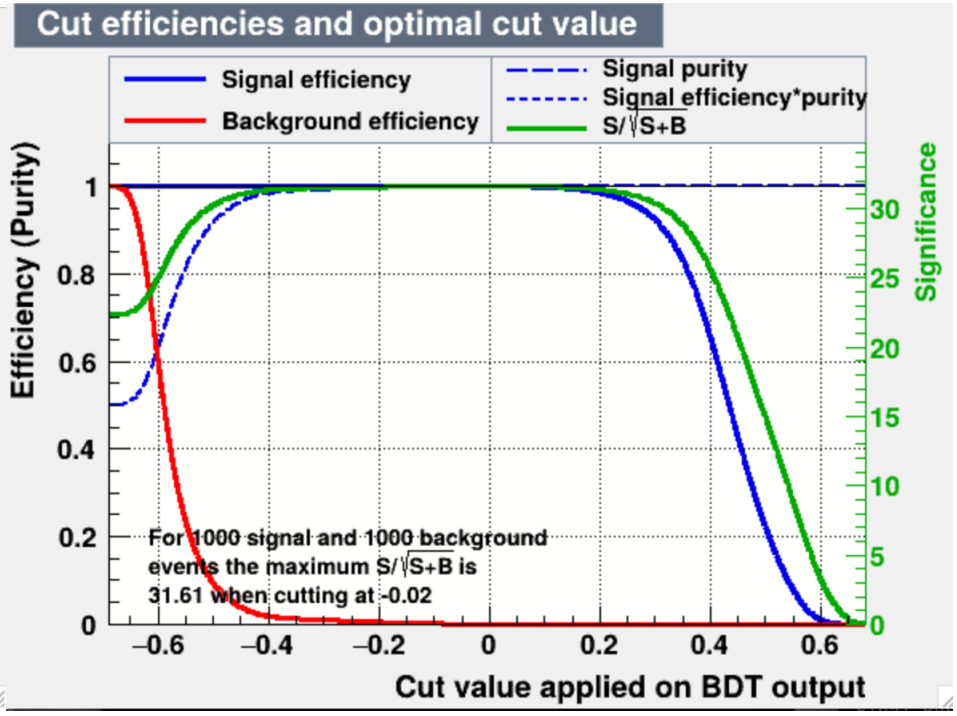
- $\texttt{Bin Error}_{\mathcal{S}} = \sqrt{\mathcal{N}} \times \frac{\sigma}{\mathcal{N}_{\text{total}}} \times \mathcal{L}$, for the expected number of events per $\mathcal{L} = 100 fb^{-1}$.



Boost Decision Trees(BDT)

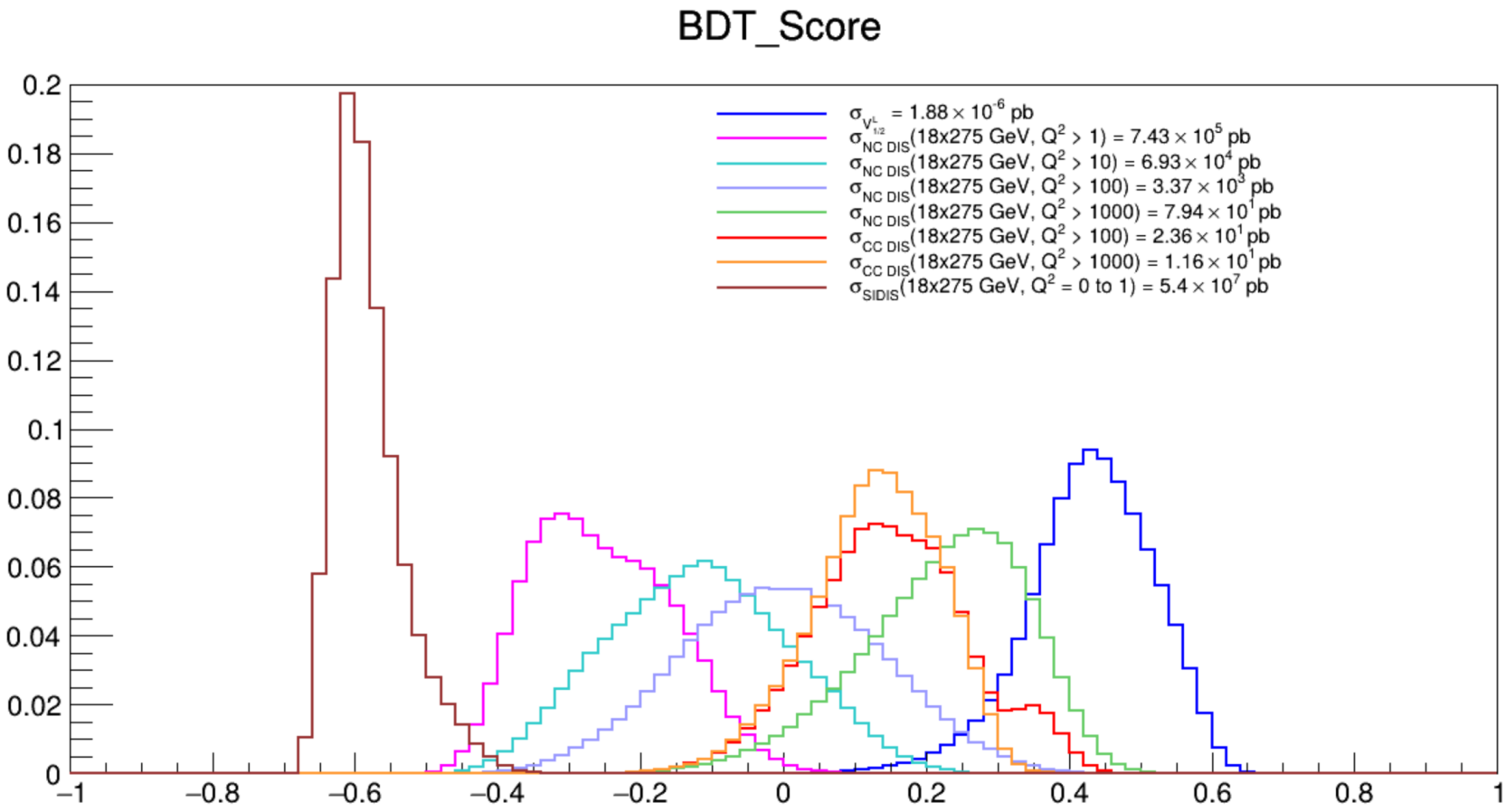
- Train the BDT Model (weights, scaling included...!)
- Signal training(500000) and testing(500000) events = 1000000.
- Background training (3500000) and testing events(3500000) = 7000000.
- Training was done with 400 Decision Tree
- dataset/weights/weights.xml



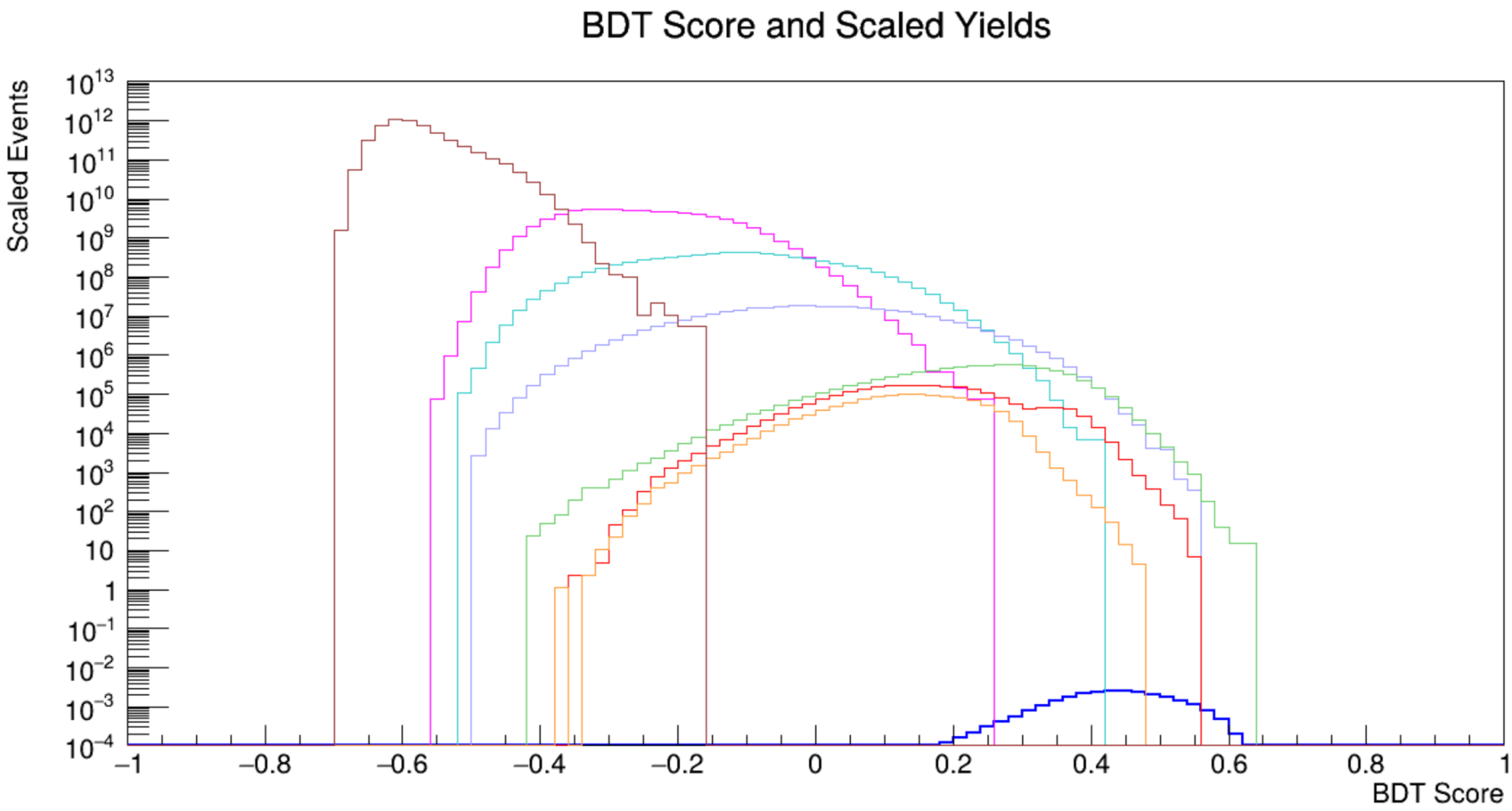


Boost Decision Trees(BDT)

- Apply the model on all other files, no need to train them all. BDT Model made it clear already what is S and B.



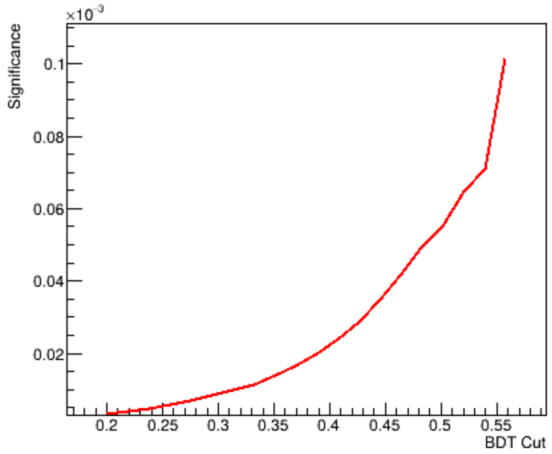
Boost Decision Trees(BDT)

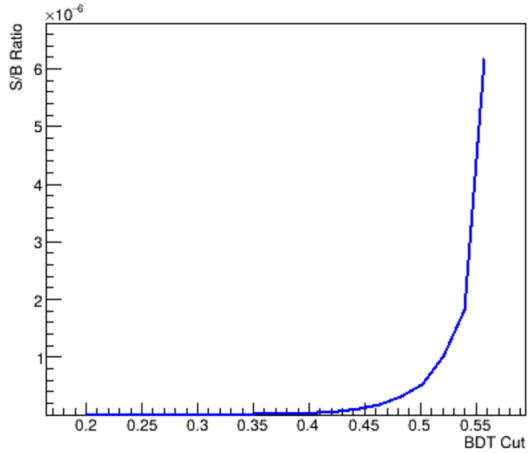


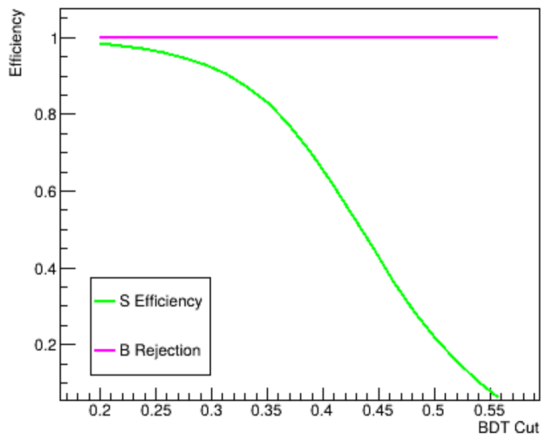
$$B = 7.43\times10^{10}+6.93\times10^{9}+3.37\times10^{8}+7.94\times10^{6}+2.36\times10^{6}+1.16\times10^{6}+5.4\times10^{13}$$

$$B = 5.408 \times 10^{13}$$

Boost Decision Trees(BDT)







BDT_Cut	Signal	Background	S/B	Significance
0.2	0.0263867	6.27307×10^{7}	4.20635×10^{-10}	3.33154×10^{-6}
0.407263	0.0167371	494085	3.3875×10^{-8}	2.38111×10^{-5}
0.520316	0.004157	4104.58	1.01277×10^{-6}	6.48852×10^{-5}
0.539158	0.00277916	1535.08	1.81043×10^{-6}	7.09329×10^{-5}
0.558	0.00166586	269.96	6.17077×10^{-6}	1.01388×10^{-4}

34 B events from NC DIS $Q^2 > 1000$, and 62159 S events survived the BDT cut application. However, considering the scaling based on the weight, the cross section for signal is very small leading to small value while the total background yield gets to 270.

Boost Decision Trees(BDT)

$$S/B_{\text{before}} = \frac{S}{\text{Total Background}} = \frac{0.02688}{5.408 \times 10^{13}} \approx 4.97 \times 10^{-16}$$

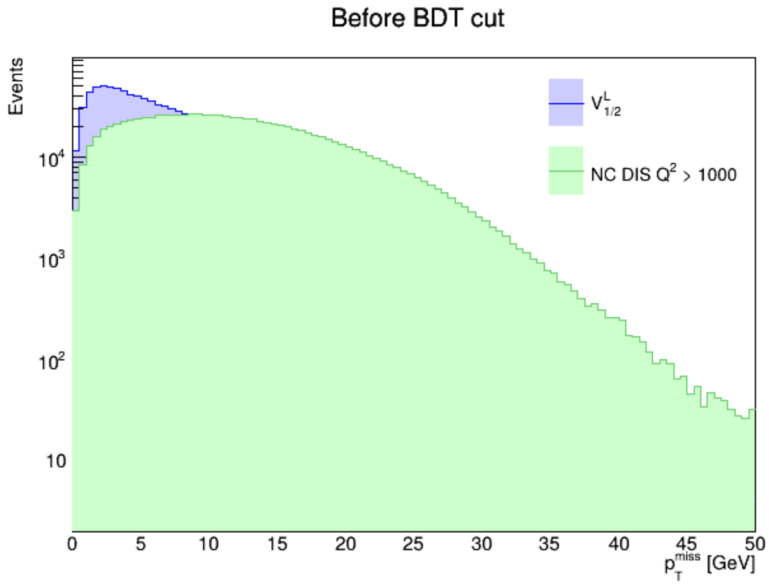
$$S/B_{\text{after}} = 6.17 \times 10^{-6}$$

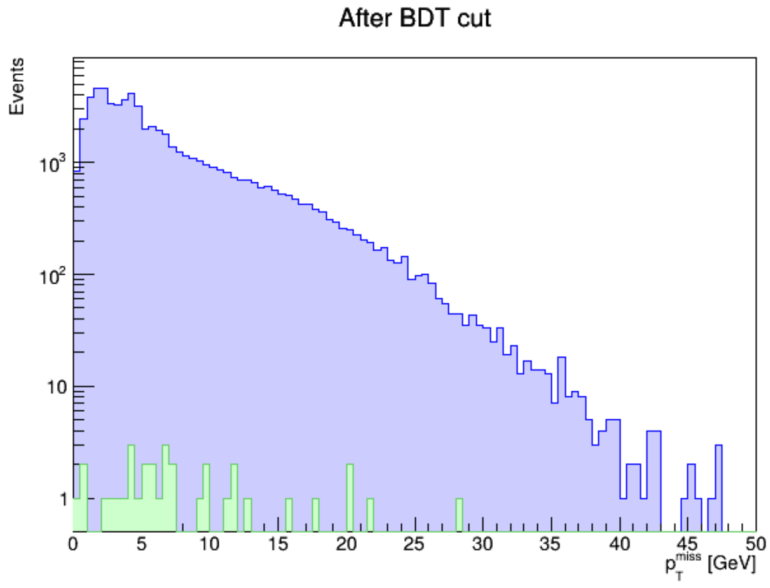
- Improvement in S/B ratio:

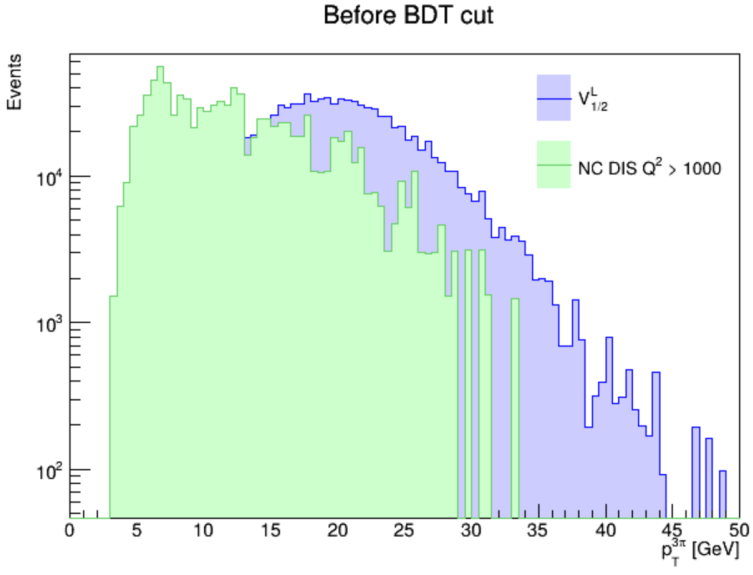
$$\text{Improvement} = 1.24 \times 10^{10}$$

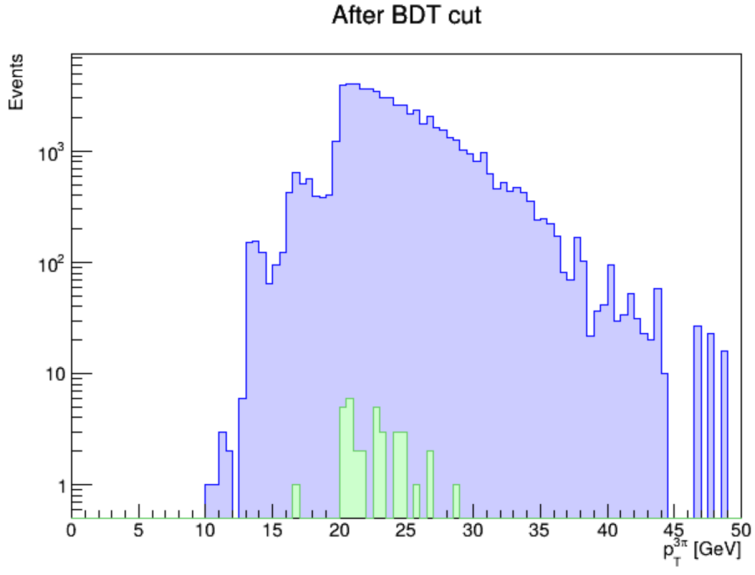
- The improvement corresponds to an enhancement of approximately **10 orders of magnitude**.

Boost Decision Trees(BDT)

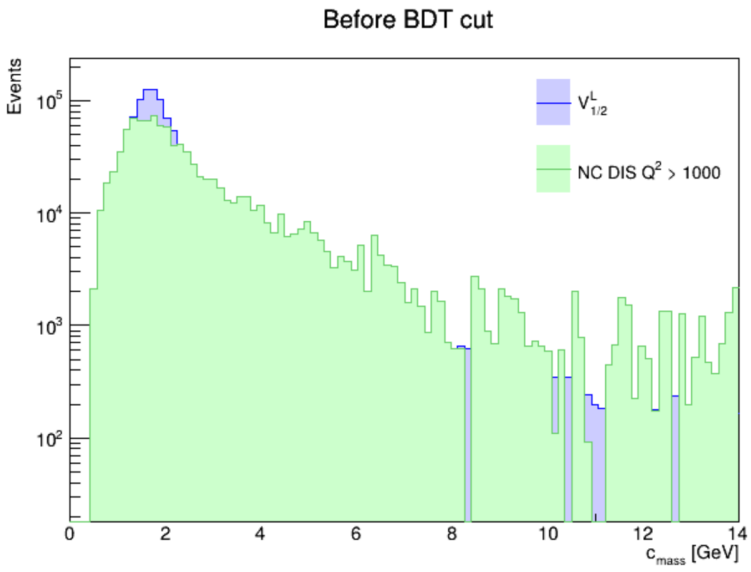


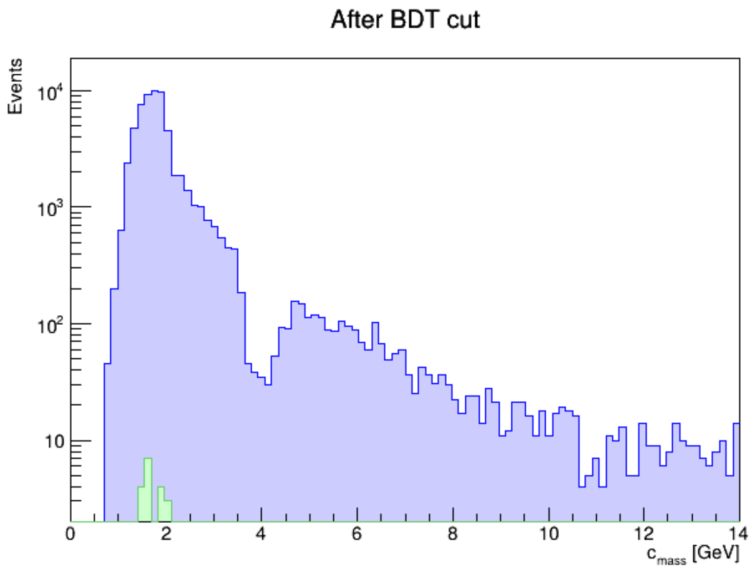






Boost Decision Trees(BDT)





Sensitivity projection for $V_{1/2}^{L}$

- Exclusion Cross-Section

$$\sigma_s = \frac{10^6}{\mathcal{L}} = \frac{10^6}{100 \text{ fb}^{-1}} = 10^4 \text{ fb} \tag{2}$$

$$\sigma_{min} = \frac{10^4 \text{ fb}}{62159} = 0.16 \text{ fb} \tag{3}$$

- Scaling for Background would yield 270 events, meaning that

$$S_{\text{min}} = 3\sqrt{270} = 49 \text{ events} \tag{4}$$

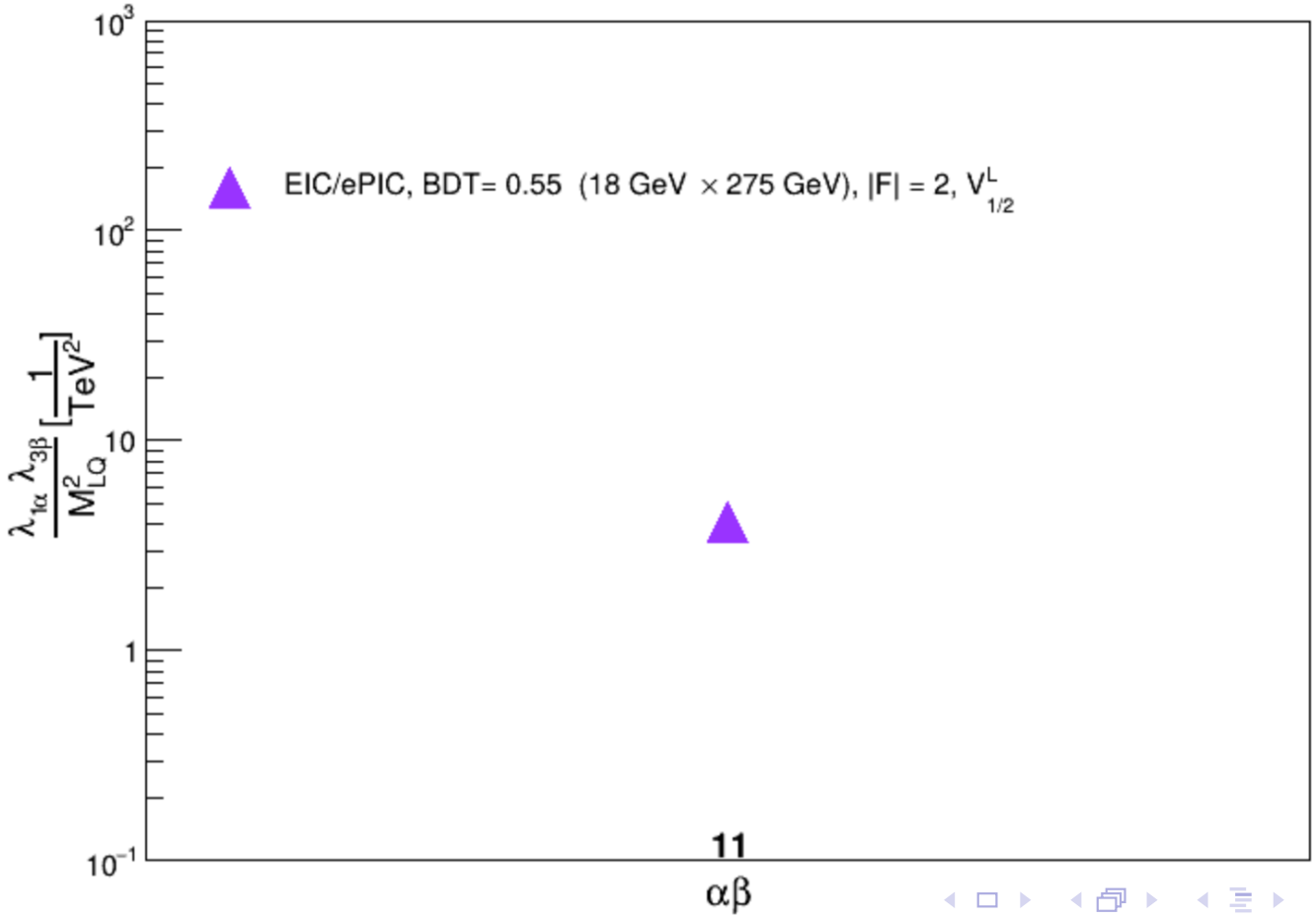
Physical interpretation: We need at least 49 signal events at 100 fb^{-1} to distinguish signal from background fluctuations at the 3σ level.

$$\boxed{\sigma_{\text{exclusion}} = 0.16 \text{ fb} \times 49 = 7.8 \text{ fb}} \tag{5}$$

Sensitivity projection for $V_{1/2}^{L}$

The exclusion boundary is represented by a **purple triangle** at position:

$$\boxed{(\alpha\beta = 11,\ y = 4.1\ \text{TeV}^{-2})} \tag{6}$$



Sensitivity projection for $V_{1/2}^{L}$

- Any leptoquark model satisfying:

$$\frac{\lambda_{11}\lambda_{31}}{M_{\mathsf{LQ}}^2} > 4.1 \text{ TeV}^{-2}$$

would have produced a detectable signal in our analysis and is therefore **excluded**.

Future work

- Project the sensitivity for other 13 LQs based on the selection criteria and BDT.
- Compare EIC results with existing limits from other experiments.
- Enhance S to B separation by feeding important input parameters to the BDT model before training it.