Opportunities for Flavour Tagging in Hadronic W decays using a Transformer NN

Freya Blekman, Florencia Canelli, Anna Macchiolo, Alexandre De Moor, Kunal Gautam, Armin Ilg, Eduardo Ploerer

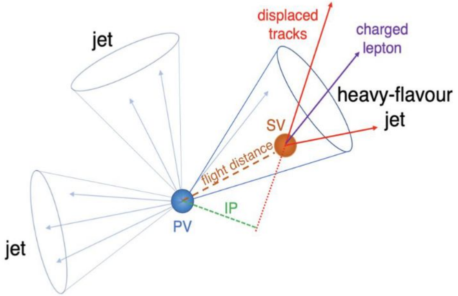
Flavour Tagging

Identification of hadronic final states is an essential ingredient in exploiting the physics potential of collider experiments

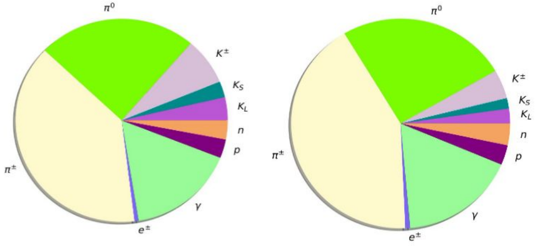
Future lepton collider such as FCC-ee offer much cleaner environment than hadronic collisions (Initial state kinematics known, no PDFs, no QCD ISR, …) => expect to do much better

Distinguishing features have historically been used to discriminate jets

- Differing colour factors for q vs g $$C_A/C_F = \frac{9}{4}$$
- Displaced SVs for b/c's
- Kaon excess for s
- Jet charge for up/down



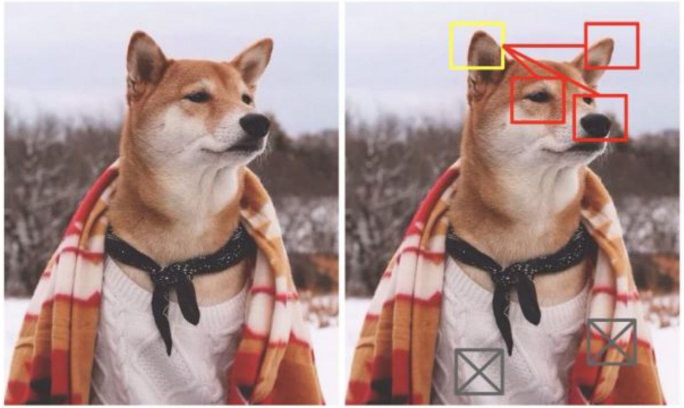
$$Q_\kappa = \frac{1}{p^\kappa_{T,jet}} \Sigma_j q_j (p^j_T)^\kappa$$



Strange jets

Down jets

Transformers



Recent advances in Natural Language Processing have come from the implementation of Attention Mechanisms

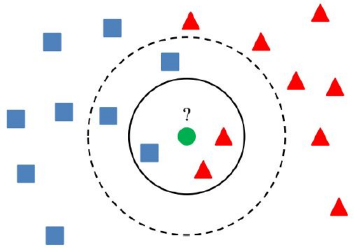
Model exploiting an attention mechanism are typically dubbed "Transformers" and some examples exist in the realm of jet flavour tagging and have achieved state-of-the-art performance

- ABCNet
- PointCloud transformer
- ParT

~200k + KNN

~2M + KNN

These architectures calculate attention on graph representations of data and can be computationally expensive

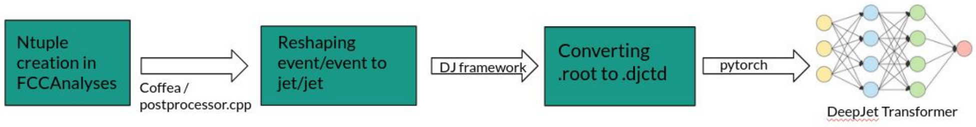


DeepJetTransformer

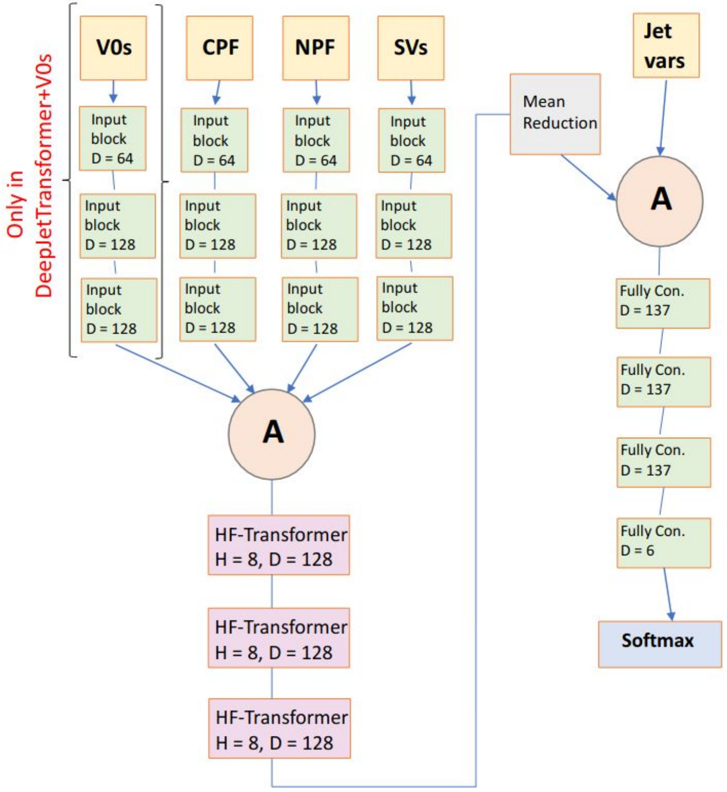
DeepJetTransformer is an architecture likewise achieving state-of-the-art performance, but using an encoder-decoder architecture

- More lightweight/still performant (~1M trainable weights, only 65k per encoder layer)

Set up pipeline for training the network using FCCAnalyses framework centrally produced FCC-ee Spring2021 samples



Code for NTuple production and for training available on github

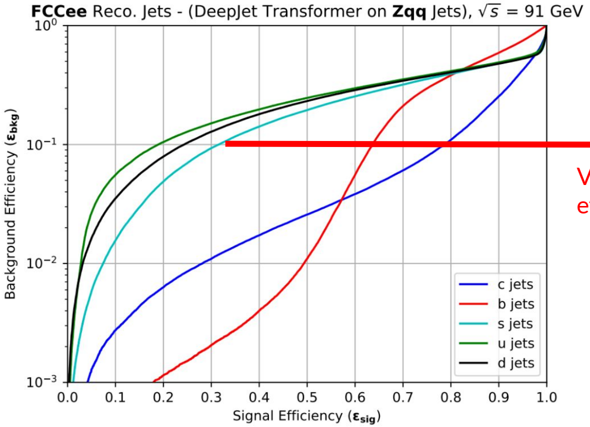


Saw our ParticleNet friends implemented ONNX inference using weaver for faster inferencing in FCCAnalyses

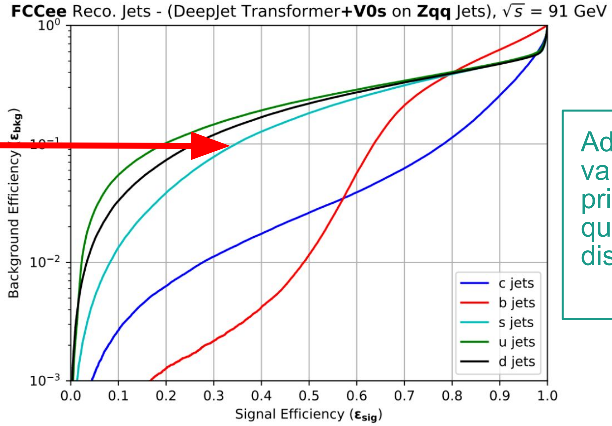
Results at the Z pole

Performance strongest for b quark jets

Weakest for up quark jets

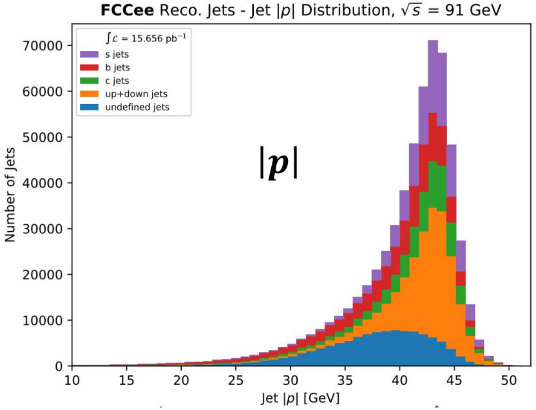


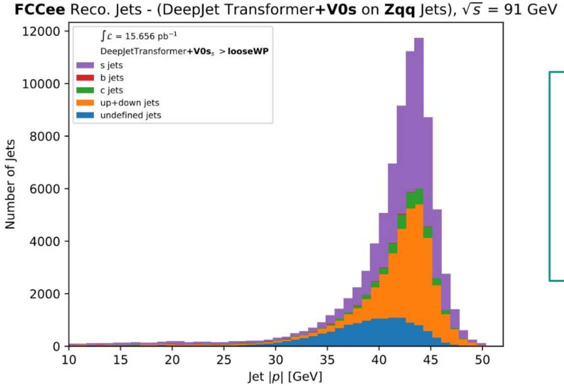
V0 improves efficiency by 3%



Addition of V0 variables improves primarily strange quark discrimination

Bkg to s quark jets almost exclusively up+down quark jets



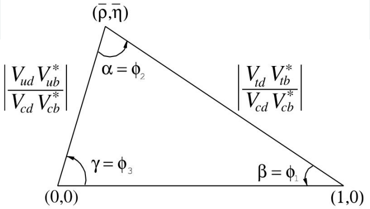


Up+down quark jet bkg relatively insensitive to V0 variables

Case Study

About to be started…

V_cs/V_cb from hadronic W decays: Motivation



Precise determinations of CKM elements offer some of the most stringent SM tests

V_cb appears in the normalization of the unitarity triangle sides, sensitivity to departure from SM will become limited by |V_cb| [1]

=> Presently most precise determinations from semileptonic B decays

$$|V_{cd}|^2+|V_{cs}|^2+|V_{cb}|^2 = 1.001 \pm 0.012 \text{ (2nd row)}$$

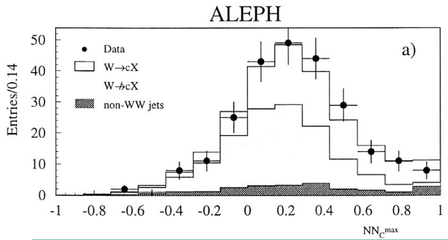
$$|V_{us}|^2+|V_{cs}|^2+|V_{ts}|^2 = 1.004 \pm 0.012 \text{ (2nd column)}$$

Precision of unitarity constraints in second row (& column) dominated by |V_cs|

Leading parametric uncertainty for 0.1% extraction of alpha_s from hadronic W decays

=> currently best determinations from D_(s) decays

V_cs in particular was extracted directly at LEP2 using hadronically decaying W->cs [2, 3, 4] achieving

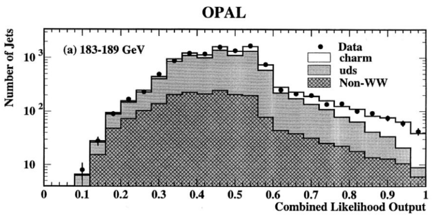


W->cX decays selected via charm tag

NN w/ 12 inputs to tag charm jets

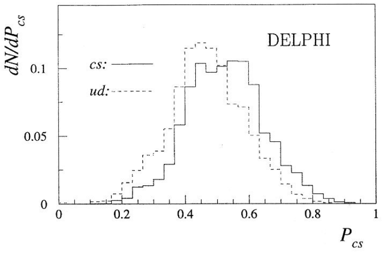
$$|V_{cs}| = 1.00 \pm 0.11_{stat} \pm 0.07_{syst}$$

$$R_c^W = \frac{|V_{cd}|^2+|V_{cs}|^2+|V_{cb}|^2}{|V_{ud}|^2+|V_{us}|^2+|V_{ub}|^2+|V_{cd}|^2+|V_{cs}|^2+|V_{cb}|^2}$$



ANN w/ 9 inputs to tag charm decay prod.

$$|V_{cs}| = 0.969 \pm 0.058$$



W->cs decays selected with charm and strange tag

Likelihood based discriminator (e.g. QGL in CMS)

$$|V_{cs}| = 0.94^{+0.32}_{-0.26}\text{(stat)} \pm 0.13\text{(syst)}$$

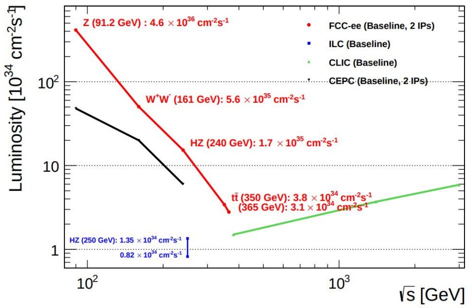
V_cs/V_cb from hadronic W decays: Outlook at FCC-ee

FCC-ee offers exquisite WW statistics at multiple E_cm

- sqrt(s) = 162 GeV ~50 10^6 WW decays (x3 10^5 LEP)
- sqrt(s) = 240 GeV ~80 10^6 WW decays (x2 10^3 LEP)
- sqrt(s) = 365 GeV ~20 10^6 WW decays

=> Immense improvement in statistical uncertainty

Alain Blondel[1], Patrick Janot[2]: FCC-ee overview: new opportunities create new challenges



Vast improvements in jet flavour tagging (like DJT) will have direct impact on precision of |V_cs|

Will enable direct extraction of |V_cb| with HF tagging

In his slides Paolo assumes the working points

- ε_b = 40% with ε_c = 10^-3, ε_uds = 10^-5
- ε_c = 60% with ε_b = 0.1, ε_uds = 0.2

⇒ **Δ** $|V_{cb}|$ (stat) (FCCee) → 0.2%(rel)

But these tagging efficiencies are an optimistic take on older results (CMS 2016), what efficiencies might one expect with DJT?

Goal → Perform realistic FullSim studies, and study impact of different detector configurations on physics potential

b/c+s-tagging : Detector Requirements

- b/c-tagging:
  - Accurate reconstruction of the decay chain, i.e. good SV reconstruction to remove background from *b/c*-jets
  - V^0 rejection to remove *s*-jet background
  - Need excellent resolution for VXD & tracker
- s-tagging:
  - Background from c-jets: requires good vertex reconstruction
  - Background from light jets: requires an excellent PID strategy to significantly improve tagging performance
  - V^0 reconstruction to identify K^0_S and Λ^0
  - Cluster counting + ToF seem to provide a good K/π separation in the momentum range of interest
- Jet clustering:
  - Irregularly shaped jets, so need a good jet definition and flavour assignment
  - Highly granular calorimeters for efficient jet reconstruction

Summary

Look at a Transformer based jet flavour tagger DeepJetTransformer in context of FCC-ee

- Lightweight (relatively) yet performant
- Trained & evaluated on Spring2021 samples
- Further training with different processes underway
- Inclusion of PID variables/estimates to address s vs u/d
- Updated results with Z flavour used for jet assignment being cross-checked

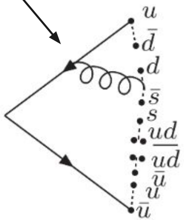
Case Studies ($W \rightarrow cs/cb$) getting started

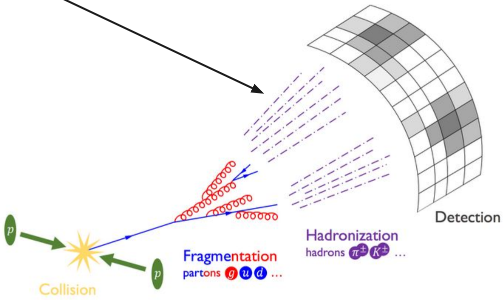
Looking forward to your ideas and interest in collaboration

Backup

The Feedback Loop

ML approaches uniquely suited for task where <u>underlying dynamics</u> poorly understood, but there is an abundance of <u>training data</u>





Goal: Study the dependence on detector requirements of physics potential

$$\text{Sensor perf.} \overset{\text{detector}}{\underset{\text{sim.}}{\rightarrow}} \text{vertexing perf.} \overset{\text{sample}}{\underset{\text{analysis}}{\rightarrow}} \text{physics perf.} \overset{\text{theory}}{\underset{\text{input}}{\rightarrow}} \text{sensor specification}$$

SVs & V⁰s for FCC

Implementation of the vertexing module from LCFIPlus

V^0 Reconstruction

- Three processes considered: $K^0_S \to \pi^+\pi^-$; $\Lambda^0 \to p\pi^-$; $\gamma_{conv} \to e^+e^-$
- Fit all possible pairs of tracks with opposite charge and constrain their invariant mass, distance from the interaction point and direction

SV Reconstruction

1. Find primary vertex and remove all tracks forming the PV
2. Remove tracks forming any V^0s
3. Find a two-track vertex (seed) that by applying a set of constraints
4. Add tracks to the seed until the resulting vertex doesn't pass the set of constraints
5. Repeat step 4&5 until no more seeds can be found