Charm Jet Tagging in Charged-Current Interactions at EIC: Tagging Optimization and Particle ID

Stephen Sekula[1] Jared Burleson[1] Miguel Arratia[2] Yulia Furletova[3]
Tim Hobbs[1,4] Olek Kusina[5] Pavel Nadolsky[1] Jae Nam[6] Fred Olness[1]
Bernd Surrow[6]

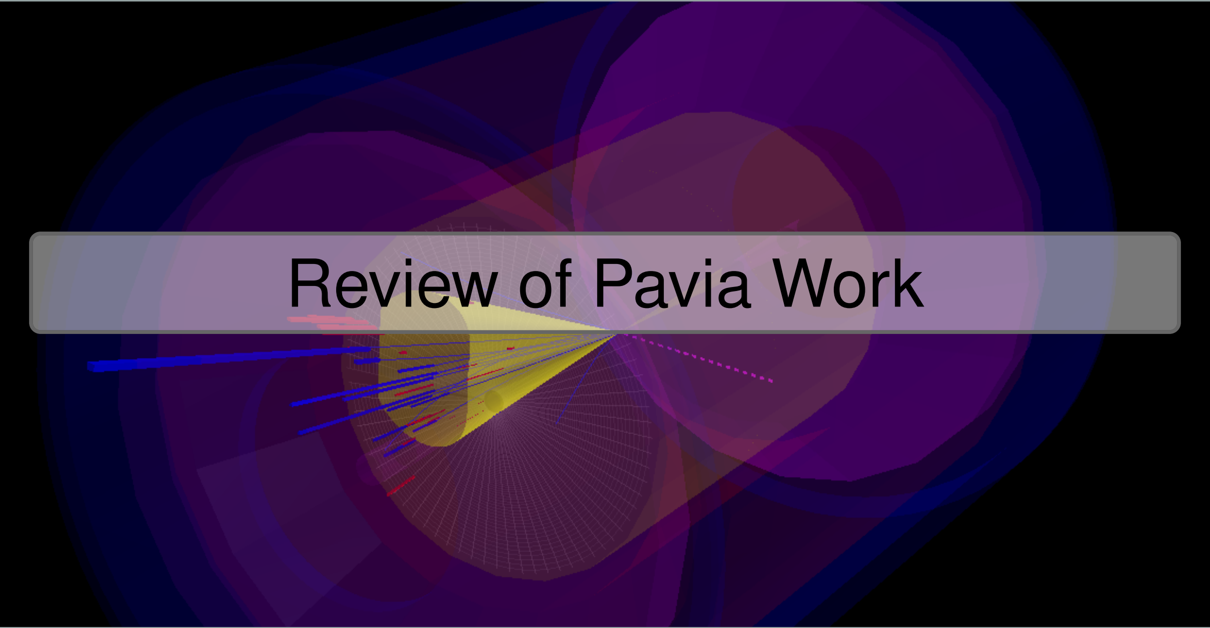
[1]Southern Methodist University, Dallas, TX; [2]University of California, Riverside, Riverside, CA; [3]Thomas Jefferson National Accelerator Laboratory, Newport News, VA; [4]EIC Center at Jefferson Lab, Newport News, VA; [5]Institute of Nuclear Physics PAN, Krakow, Poland [6]Temple University, Philadelphia, PA

June 8, 2020

SMU | DEDMAN COLLEGE OF HUMANITIES & SCIENCES   UC RIVERSIDE UNIVERSITY OF CALIFORNIA   Jefferson Lab   IFJ PAN

Outline

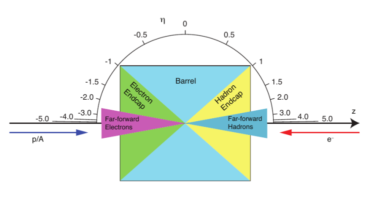
Review of Pavia Work



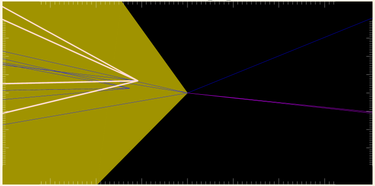
Review of the Pavia Work

- Tracking: covers $|\eta| < 3.5$ with 85%-98% efficiency depending on p_T and η; track impact parameter resolution is $20\mu\mathrm{m}$ in d_0 and z_0. Tracking system immersed in 1.5T magnetic field.
- ECal covers $|\eta| < 4$ with resolution of $\sigma = E \times (1.0\%) \oplus \sqrt{E} \times (2.0\%)$ in the backward direction worsening to $E \times (2.0\%) \oplus \sqrt{E} \times (12.0\%)$ in the forward direction. Granularity is $(\Delta\eta, \Delta\phi) = (0.020, 0.020)$.
- HCal covers $|\eta| < 4$ with resolution of $\sigma = E \times (10\%) \oplus \sqrt{E} \times (50\%)$ in the backward and forward direction, worsening to $E \times (10.0\%) \oplus \sqrt{E} \times (100\%)$ in the barrel. Granularity is best in forward/backward region, $(\Delta\eta, \Delta\phi) = (0.025, 0.025)$, and more coarse in the barrel, $(\Delta\eta, \Delta\phi) = (0.1, 0.1)$

We used signed, high-impact parameter track counting as a means of tagging anti-k_T jets (R=1). We found an average 5% charm-jet efficiency and 0.02% light-jet efficiency, with no formal optimization.

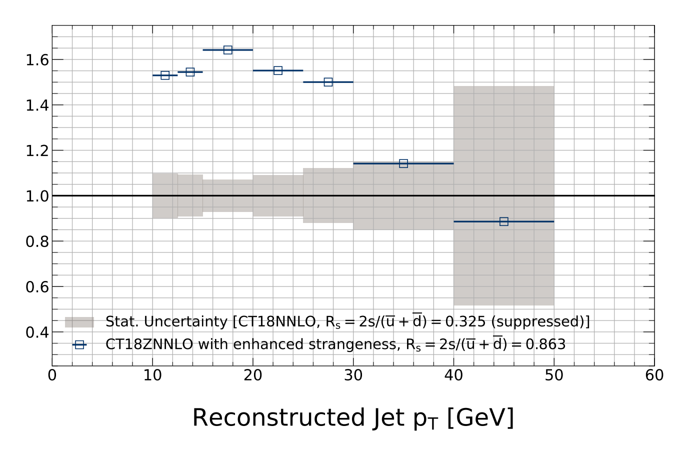


PYTHIA8 + CT18NNLO PDFs for collider simulation (unpolarized 10x275 configuration); detector response modeled using Delphes [1][2] based on EIC Detector Matrix [3].



Projected Uncertainty on Charm Jet Yield in 100 fb^{-1}

One key message for Pavia was the estimate of the uncertainty on the charm jet yield in CC DIS with 100 fb^{-1} of data. Using the CT18 PDFs, including two variations (suppressed and enhanced strangeness) representing of the range of tension in proton strangeness from existing PDF fits, we suggested this approach could help resolve this tension → especially as part of an eventual global EIC analysis.



Progress and Expansion Since Pavia

- The presentation at Pavia generated a lot of useful discussion during and after the meeting, and our group has expanded in number and scope since then.
- We're very excited to be working together now with a few new people:
  - Bernd Surrow and Jae Nam (Temple University) are pursuing studies of vertexing, secondary vertex studies, etc. These will be potent flavor-tagging tools for heavy flavor jets.
  - Jared Burleson (SMU) joins as a summer research student; his existing experience in using jet substructure to study Higgs decays to gluons will benefit flavor-tagging efforts with large-radius (R=1) heavy flavor jets.
- New studies since Pavia:
  - We did not optimize the hyper-parameters for the high-impact parameter (HIP) track-based flavor tagging $\rightarrow$ we've improved this since then to see if optimization can compensate for detector performance scenario changes.
  - We did not look at dedicated particle ID (PID) as a component of flavor tagging $\rightarrow$ we've begun a first round of studies on this.

Updated High-Impact Parameter Tagging Work

NEW: Optimization Procedure

This work is all performed on a 10x275 EIC unpolarized electron-proton collision scheme. There are a few hyperparameters for the track-counting flavor-tagging approach:

- The minimum number of “high-impact-parameter” (HIP) tracks;
- The minimum flight significance (3-D) to define a track as “HIP”;
- The minimum p_{T} of tracks to be considered for HIP designation.

I have selected the following figure of merit (FOM) for hyperparameter optimization, which served me well on the BaBar and ATLAS experiments for first-stage analysis development: the Punzi FOM [**?**]:

$$\mathcal{F} = \frac{\varepsilon_s}{\frac{N_\sigma}{2} + \sqrt{N_b}}$$

where ε_s is the efficiency of selecting signal; σ is the target significance of separating signal from background; and N_b is the expected number of background events ($\mathcal{L} \times \sigma_{\mathrm{CC-DIS}} \times \varepsilon_b$). I use $N_\sigma = 5$.

Optimization Grid

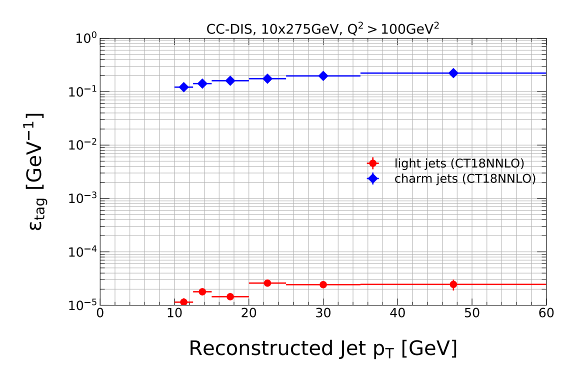
With the goal of optimizing the separation of light and charm jets, I scan the hyperparameters as follows:

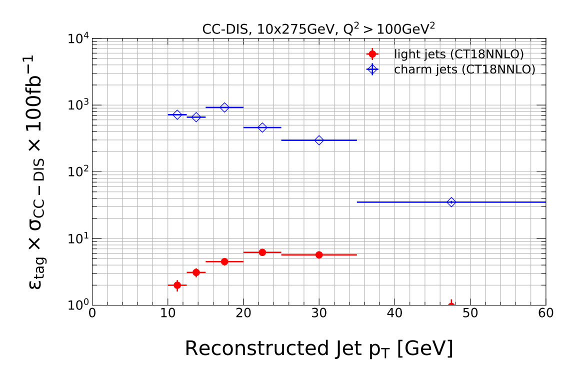
- N_{HIP}^{min}: 2, 3, 4
- sIP_{3D}: 1.5-4.0 in steps of 0.25; then 5.0-9.0 in unit steps.
- p_{T}^{track}: 0.25-1.5 in steps of 0.25

The maximum FOM is determined to be 0.016 for $N_{HIP}^{min} = 2$, $\mathrm{sIP}_{3D} > 3.75$, and $p_{T}^{track} > 0.75\ \mathrm{GeV}$, which yields an expected charm-jet tagging efficiency of 13% (for jets with $p_T > 5\ \mathrm{GeV}$), a light-jet efficiency of 2×10^{-5} for a target sample of $100\,\mathrm{fb}^{-1}$; this yields an expectation of about 4000 charm-jet CC-DIS events in that same sample.

Compare this to our unoptimized 5% charm-jet efficiency (20×10^{-5} light-jet efficiency) from Pavia.

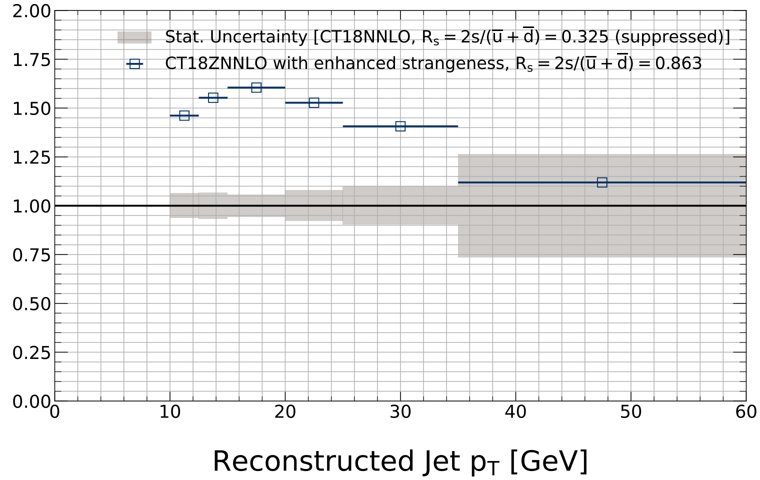
Efficiency and Yield Plots





Compared to the previous hyperparameter choices, with no real optimization applied, this is a vast improvement in performance.

Updated projected statistical uncertainty in 100 fb^{-1} compared to variation in strangeness suppression ratio:



A First Look at Track Impact Parameter Resolution Effects

Track impact parameter resolution plays a key part in the HIP track-tagging approach. With a baseline optimization framework in place, it is useful to study the effect on tagging, with re-optimization, of the degradation in track impact parameter resolution.

		Efficiency [%]		
σ_{d_0} (μm)	σ_{z_0} (μm)	Charm Jet (%)	Light Jet (%)	% Change
20	20	12.6	2×10^{-3}	NA
100	100	8.0	2×10^{-3}	-32%
20	100	10.2	2×10^{-3}	-19%
10	10	13.3	2×10^{-3}	+5.6%

This preliminary work shows that such large, across-the-board degradations in one or both impact parameter resolutions will result in real charm-jet losses even while re-optimizing and holding the light-jet efficiency constant. This will be useful input to future studies on more realistic “progressive” losses in resolution with η, etc. Improving (degrading) resolution leads the optimization to loosen (tighten) the cut on $\mathrm{sIP_{3D}}$.

Particle Identification

We employ a basic but flexible particle ID approximation in the analysis code.

- Electrons
  - Tracks are searched for true electrons and true pions. If a track is a true electron, I select it as a "reconstructed electron" assuming 90% identification efficiency. If a track is a true pion, I assume 2.4σ separation between electrons and pions in the PID algorithm (a 2% pion mis-ID rate, assuming Gaussian probabilities). This model is based on an ECAL-only PID approach.
- Kaons
  - Tracks are searched for true kaons and true pions. If a track is a true kaon, I select it as a "reconstructed kaon" assuming 90% identification efficiency. If a track is a true pion, I assume 3σ separation between kaons and pions in the PID algorithm (a 0.44% pion mis-ID rate, assuming Gaussian probabilities).
- Muons
  - Tracks are searched for true muons and true pions. If a track is a true muon, I select it as a "reconstructed muon" assuming 95% identification efficiency. If a track is a true pion, I assume 2σ separation between muons and pions in the PID algorithm (a 5.4% pion mis-ID rate, assuming Gaussian probabilities).

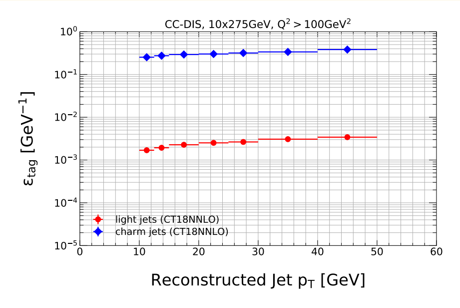
Pions mis-identified at the previous stage are removed for consideration in the next stage ($e \rightarrow K \rightarrow \mu$)

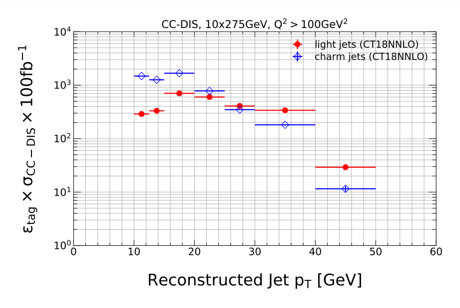
A Quick Look at the PID content of jets untagged by the $\mathrm{sIP_{3D}}$ Approach

These observations are based on a sample of 20M CC DIS events. A charm jet is electron- (or muon- or kaon-) tagged if it contains at least 1 electron with $p_{T}^{track} > 1.0\ \mathrm{GeV}$ whose impact parameter 3-D displacement significance is $\geq 3\sigma$. Yields in this table are not weighted to the target luminosity or cross-section of CC DIS.

Criteria	Light Jets (% of total)	Charm Jets (% of total)
None	19724681	438412
$\mathrm{sIP_{3D}}$-tagged	386 (0.002%)	55324 (12.6%)
$\mathrm{sIP_{3D}}$-untagged	19724295 (99.998%)	383088 (87.4%)
$\mathrm{sIP_{3D}}$-untagged and electron-tagged	5022 (0.025%)	10104 (2.3%)
$\mathrm{sIP_{3D}}$-untagged and muon-tagged	9670 (0.049%)	13288 (3.0%)
$\mathrm{sIP_{3D}}$-untagged and kaon-tagged	27451 (0.139%)	36092 (8.2%)
$\mathrm{sIP_{3D}}$-untagged and $(e/\mu/K)$-tagged	42120 (0.214%)	58711 (13.4%)
Tagged by any means	42506 (0.22%)	114035 (26.0%)

Some Plots - Efficiency and Yield in $100\,\mathrm{fb}^{-1}$





While this first look at single-particle taggers holds some promise, it's also clear that light jets more easily fake such PID-only tagged charm jets. This looks like another good opportunity for a pseudo-optimization of each single-particle tagger and/or the pursuit of a basic multivariate tagger that can combine discriminant information to try to improve signal-to-noise.

Conclusions and Outlook

Conclusions and Outlook

- We've had a first look at optimization of the signed-HIP tagging approach.
  - It's possible to achieve something like 13% tagging efficiency with this approach. However, altering the track impact-parameter resolution has a strong effect on performance, even with optimization. Degrading the resolution from 20μm can drop performance by 20-30%; increasing the resolution can improve performance.
- We've started looking at PID as an additional means to gain sensitivity.
  - Our baseline PID scenarios are consistent with expectations from modern, common detector technologies/approaches/choices. Single-track PID brings much larger backgrounds, and needs optimization. It is likely a healthy part of a multivariate strategy for tagging.
- Future directions
  - Vertexing and secondary vertexing
  - Jet substructure in large-radius jets
  - Further alterations to the detector scenario, as motivated by interests and needs for the EIC program

References I

[1] https://github.com/miguelignacio/delphes_EIC.

[2] https://github.com/stephensekula/delphes_EIC.

[3] "EIC Detector Matrix," 2020. https://physdiv.jlab.org/DetectorMatrix/.