Integrating EFNs with Jet Substructure Observables for Enhanced Jet Quenching Studies

ML4JETS 2024
Paris, France

João A. Gonçalves
jgoncalves@lip.pt
Supervised by:
Guilherme Milhano

LIP - Lisboa
IST - ULisboa

November 5, 2024

Content

Outline

Goal:

Goal: Discriminate vacuum-like jets from quenched jets.

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy:

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy: Discriminate between pp jets (all vacuum-like)
and PbPb jets (mix of vacuum-like and quenched jets).

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy: Discriminate between pp jets (all vacuum-like)
and PbPb jets (mix of vacuum-like and quenched jets).

Challenges:

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy: Discriminate between pp jets (all vacuum-like)
and PbPb jets (mix of vacuum-like and quenched jets).

Challenges:

1. Medium Response (MR) aids models in identifying PbPb jets.

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy: Discriminate between pp jets (all vacuum-like) and PbPb jets (mix of vacuum-like and quenched jets).

Challenges:

1. Medium Response (MR) aids models in identifying PbPb jets.
2. Underlying Event (UE) contamination degrades discrimination power to levels similar to those without MR.

Introduction
QGP

Goal: Discriminate vacuum-like jets from quenched jets.

Proxy: Discriminate between pp jets (all vacuum-like)
and PbPb jets (mix of vacuum-like and quenched jets).

Challenges:

1. Medium Response (MR) aids models in identifying PbPb jets.
2. Underlying Event (UE) contamination degrades discrimination power to levels similar to those without MR.
3. Only by considering all these effects can we approach a physically meaningful result.

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

[6] doi.org/10.1007/JHEP04(2018)013

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

$$\bullet_j \iff \sum_{i_j=1}^{M} z_{i_j}$$

[6] doi.org/10.1007/JHEP04(2018)013

$$\text{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

$$\bullet_j \iff \sum_{i_j=1}^{M} z_{i_j}$$

$$k \text{ ———— } \ell \iff \theta_{i_k i_\ell}$$

[6] doi.org/10.1007/JHEP04(2018)013

Introduction
Observables

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}, \qquad \bullet_j \iff \sum_{i_j=1}^{M} z_{i_j} \qquad k \text{ ——— } \ell \iff \theta_{i_k i_\ell}$$

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_1 i_3} \theta_{i_1 i_4} \theta_{i_1 i_5} \theta_{i_4 i_5}^2$$

[6] doi.org/10.1007/JHEP04(2018)013

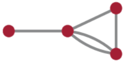
Introduction
Observables

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

$$\bullet_j \iff \sum_{i_j=1}^{M} z_{i_j}$$

$$k \;\text{———}\; \ell \iff \theta_{i_k i_\ell}$$

$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta^2_{i_4 i_5}$$



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta^2_{i_2 i_4}\theta_{i_3 i_4}$$

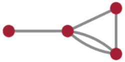
[6] doi.org/10.1007/JHEP04(2018)013

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

 $_j \iff \sum_{i_j=1}^{M} z_{i_j}$

k  $\ell \iff \theta_{i_k i_\ell}$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta_{i_4 i_5}^2$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_2 i_4}^2\theta_{i_3 i_4}$

$$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2}\left(\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta\phi_{i_1 i_2})\right) = \frac{1}{2} \times$$  $+ \cdots$

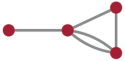
[6] doi.org/10.1007/JHEP04(2018)013

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

 $_j \iff \sum_{i_j=1}^{M} z_{i_j}$

k  $\ell \iff \theta_{i_k i_\ell}$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta_{i_4 i_5}^2$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_2 i_4}^2\theta_{i_3 i_4}$

$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta\phi_{i_1 i_2})) = \frac{1}{2} \times$  $+ \cdots$

$\lambda^{(2)} = \frac{1}{2} \sum_{i\in J}\sum_{j\in J} z_i z_j \theta_{ij}^2 = \frac{1}{2} \times$ 

[6] doi.org/10.1007/JHEP04(2018)013

Introduction

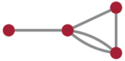
Observables

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

 $_j \Longleftrightarrow \sum_{i_j=1}^{M} z_{i_j}$

k  $\ell \Longleftrightarrow \theta_{i_k i_\ell}$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1}z_{i_2}z_{i_3}z_{i_4}z_{i_5}\theta_{i_1i_2}\theta_{i_2i_3}\theta_{i_1i_3}\theta_{i_1i_4}\theta_{i_1i_5}\theta^2_{i_4i_5}$

 $= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1}z_{i_2}z_{i_3}z_{i_4}\theta_{i_1i_2}\theta_{i_2i_3}\theta^2_{i_2i_4}\theta_{i_3i_4}$

$$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1}z_{i_2}\left(\cosh(\Delta y_{i_1i_2}) - \cos(\Delta\phi_{i_1i_2})\right) = \frac{1}{2} \times$$

 $+ \cdots$

$$\lambda^{(2)} = \frac{1}{2}\sum_{i\in J}\sum_{j\in J} z_i z_j \theta_{ij}^2 = \frac{1}{2} \times$$

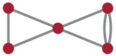
$\lambda^{(4)} =$  $- \frac{3}{4} \times$ 

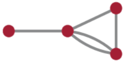
[6] doi.org/10.1007/JHEP04(2018)013

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

 $j \iff \sum_{i_j=1}^{M} z_{i_j}$

 $\iff \theta_{i_k i_\ell}$

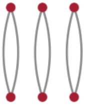
 $$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta^2_{i_4 i_5}$$

 $$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta^2_{i_2 i_4}\theta_{i_3 i_4}$$

$$\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta\phi_{i_1 i_2})) = \frac{1}{2} \times$$  $+ \cdots$

$$\lambda^{(2)} = \frac{1}{2}\sum_{i\in J}\sum_{j\in J} z_i z_j \theta^2_{ij} = \frac{1}{2} \times$$ 

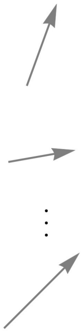
$\lambda^{(4)} =$  $- \frac{3}{4} \times$ 

$\lambda^{(6)} =$  $- \frac{3}{2} \times$  $+ \frac{5}{8} \times$ 

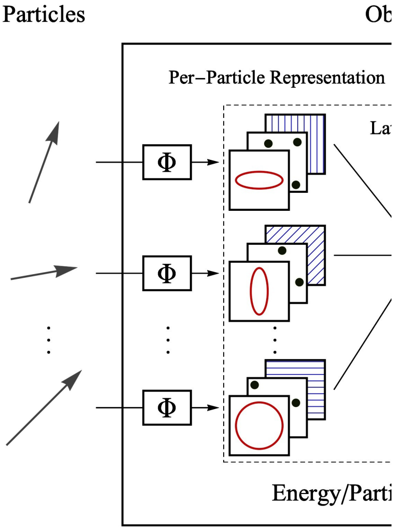
[6] doi.org/10.1007/JHEP04(2018)013

Introduction
ML

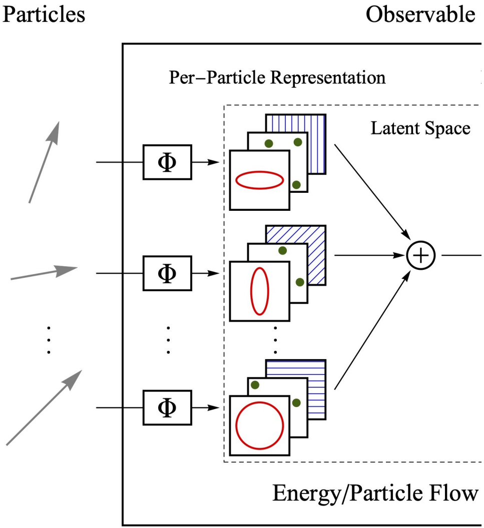
Particles



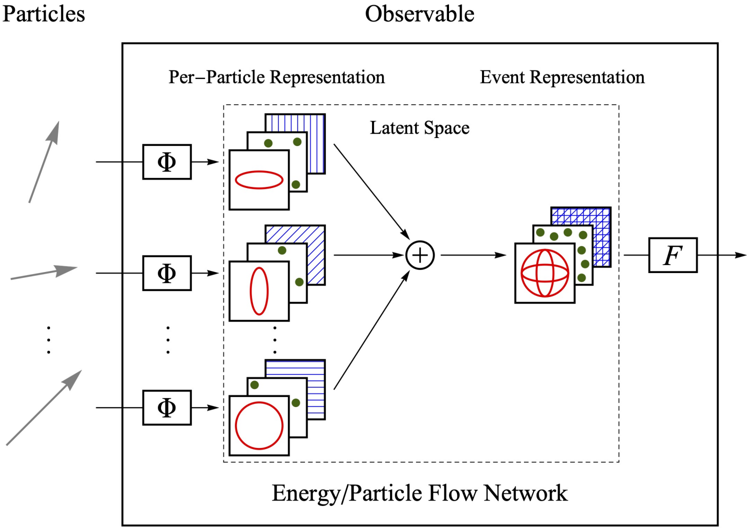
arXiv:1810.05165 [hep-ph]



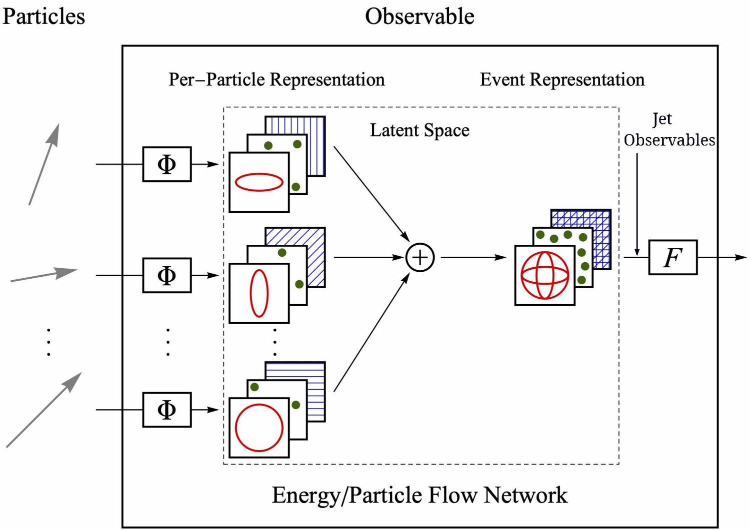
arXiv:1810.05165 [hep-ph]



arXiv:1810.05165 [hep-ph]



arXiv:1810.05165 [hep-ph]



arXiv:1810.05165 [hep-ph]

Outline

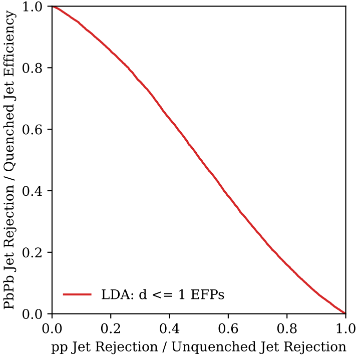
Results

Energy Flow Polynomials (LDA)

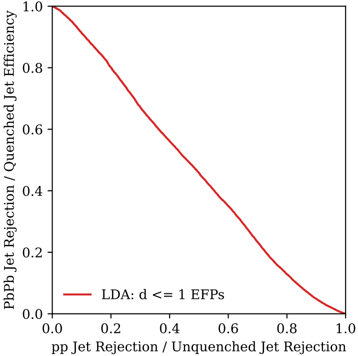
Paper in Construction

Paper in Construction

No UE



With subtracted UE

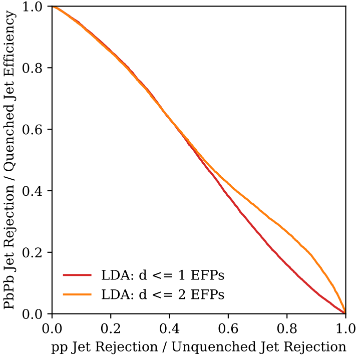


arXiv:1712.07124 [hep-ph]

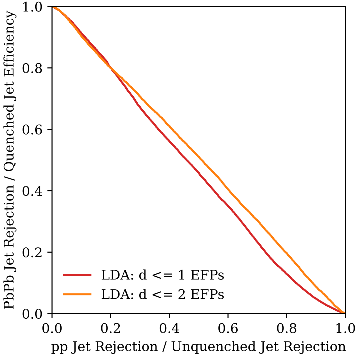
Results
Energy Flow Polynomials (LDA)

Paper in Construction

No UE



With subtracted UE



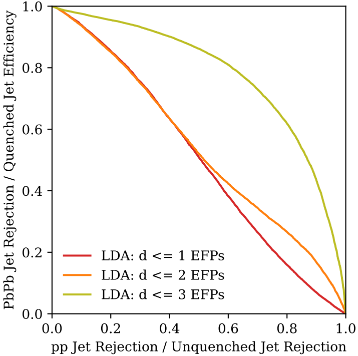
arXiv:1712.07124 [hep-ph]

Results

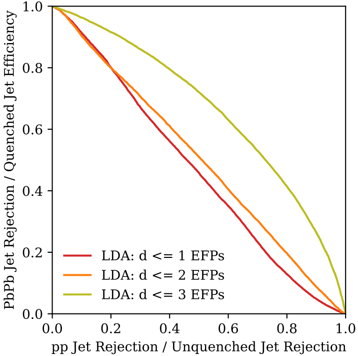
Energy Flow Polynomials (LDA)

Paper in Construction

No UE



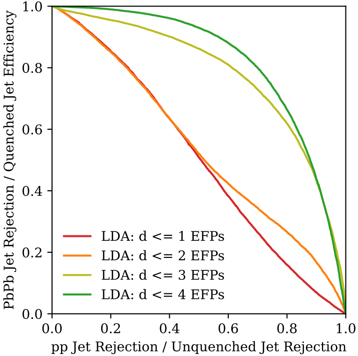
With subtracted UE



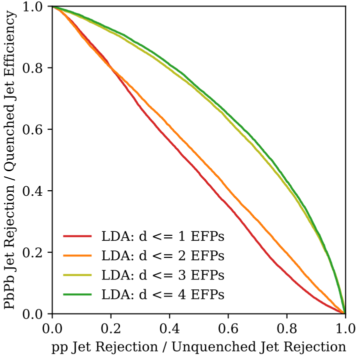
arXiv:1712.07124 [hep-ph]

Paper in Construction

No UE



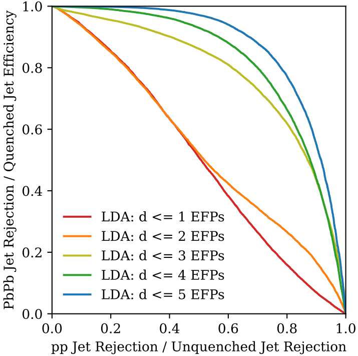
With subtracted UE



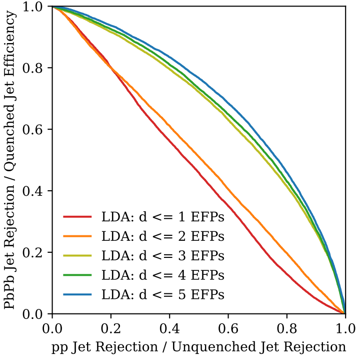
arXiv:1712.07124 [hep-ph]

Paper in Construction

No UE



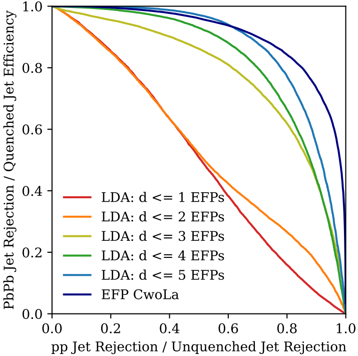
With subtracted UE



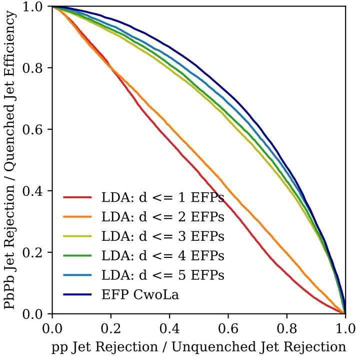
arXiv:1712.07124 [hep-ph]

Paper in Construction

No UE



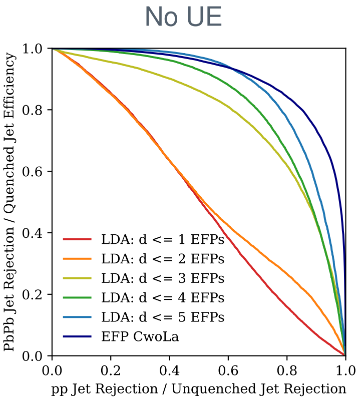
With subtracted UE

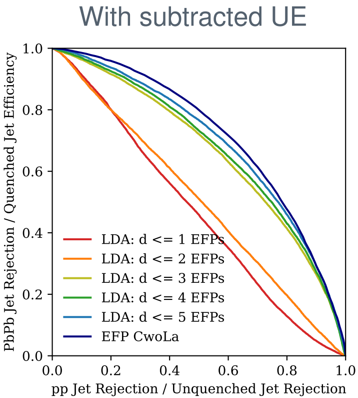


arXiv:1712.07124 [hep-ph]

Results
Energy Flow Polynomials (LDA)

Paper in Construction





Classification on gen level, picks up on the medium response and the model performs very well.

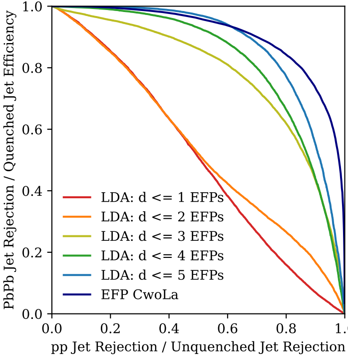
arXiv:1712.07124 [hep-ph]

Results

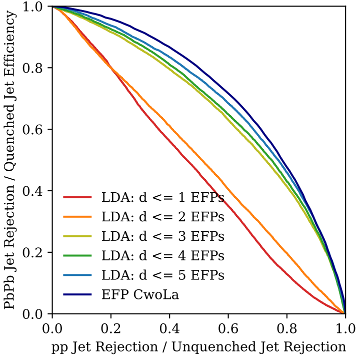
Energy Flow Polynomials (LDA)

Paper in Construction

No UE



With subtracted UE

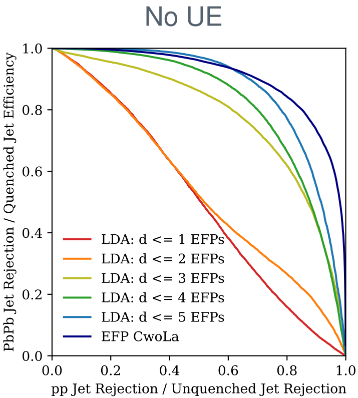


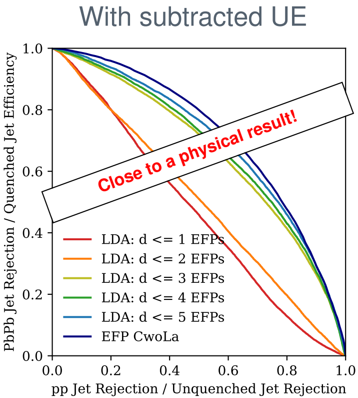
Classification on gen level, picks up on the medium response and the model performs very well.

Applying the procedure greatly reduces the discrimination power (AUC from ~ 0.8675 to ~ 0.6964).

arXiv:1712.07124 [hep-ph]

Paper in Construction





Classification on gen level, picks up on the medium response and the model performs very well.

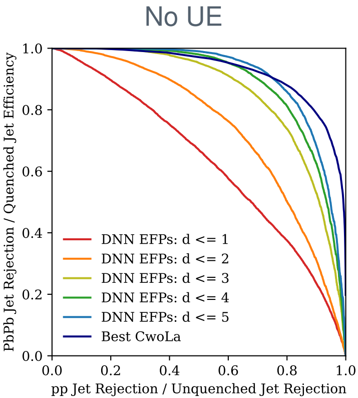
Applying the procedure greatly reduces the discrimination power (AUC from ~ 0.8675 to ~ 0.6964).

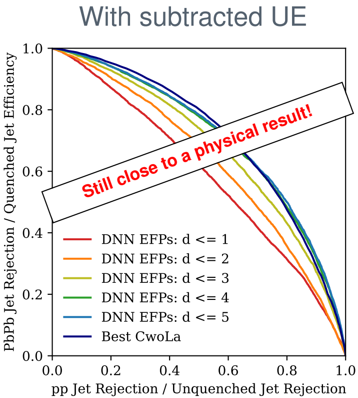
arXiv:1712.07124 [hep-ph]

Results

Energy Flow Polynomials (DNN)

Paper in Construction





Same behavior from no UE to with UE contamination.

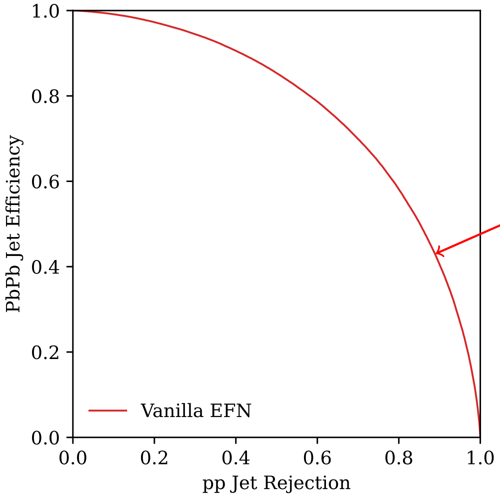
Some gain in discrimination power (AUC from ~ 0.9067 to ~ 0.7142).

arXiv:1712.07124 [hep-ph]

Results
Energy Flow Networks

Paper in Construction

Paper in Construction

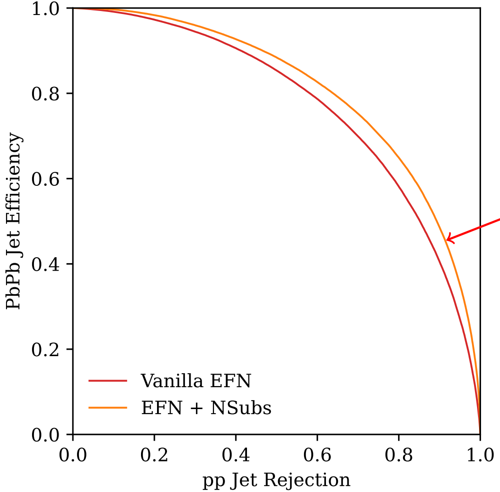


Vanilla EFN with simple architecture provides our baseline.

AUC = 0.7652 ± 0.0004

arXiv:1712.07124 [hep-ph]

Paper in Construction



Including N-Subjetiness in the F network increases performance.

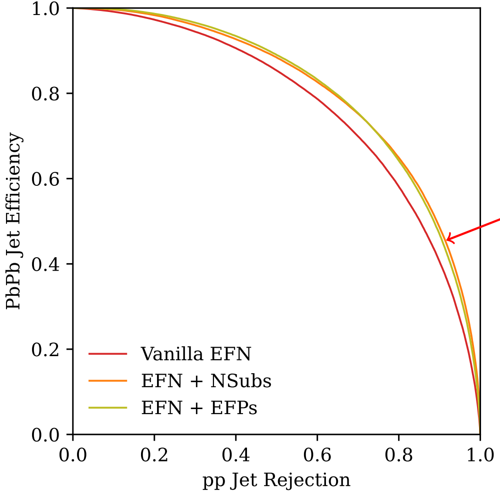
AUC = 0.8053 ± 0.0003

arXiv:1712.07124 [hep-ph]

Results

Energy Flow Networks

Paper in Construction

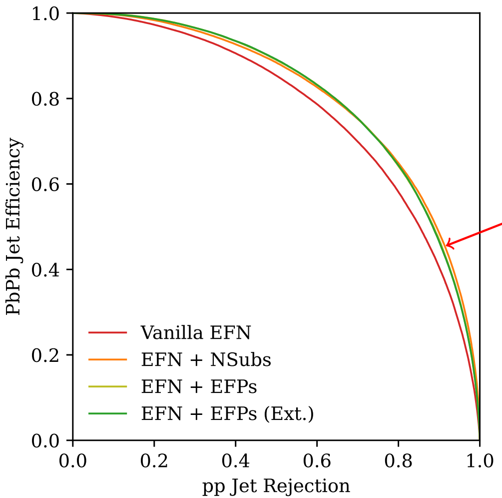


We obtain similar performance including EFPs.

AUC = 0.8050 ± 0.0004

arXiv:1712.07124 [hep-ph]

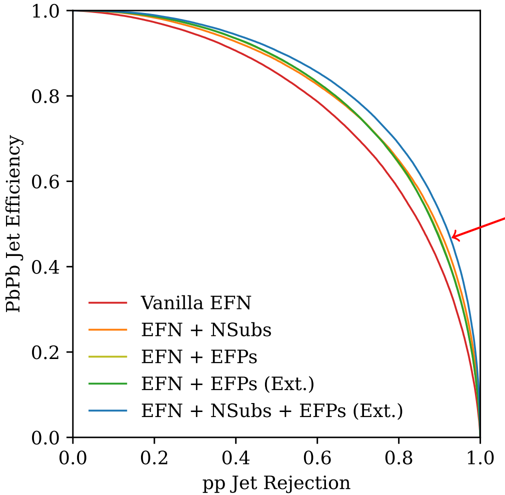
Paper in Construction



Adding EFPs with different κ and β does not seem beneficial.
Likely due to the fixed arch and the larger number of obs..
AUC = 0.8020 ± 0.0055

arXiv:1712.07124 [hep-ph]

Paper in Construction

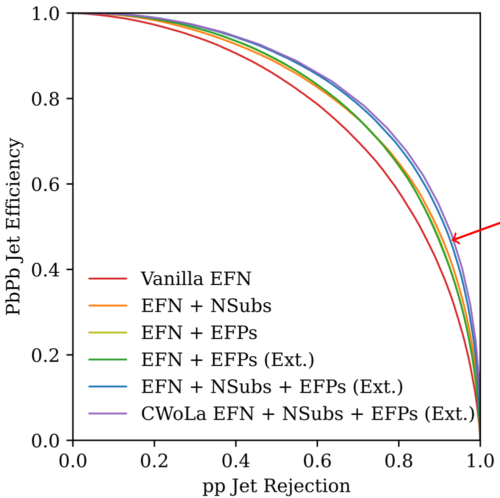


Combining all observables, results in a gain in performance.

AUC = 0.8268 ± 0.0009

arXiv:1712.07124 [hep-ph]

Paper in Construction



With CWoLa we asses "would be" Quenched discrimination.
AUC = 0.8351

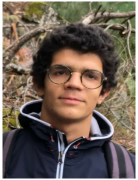
arXiv:1712.07124 [hep-ph]

Paper in Construction

Results
Energy Flow Networks with Particle Distances

Paper in Construction

By: Martim Pinto

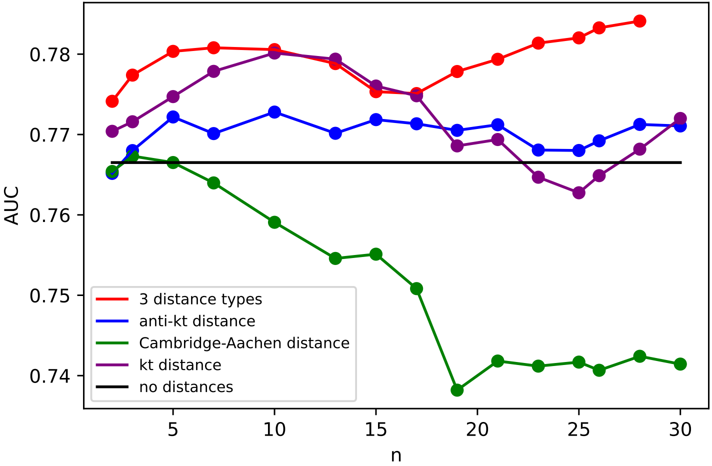


Summer Student

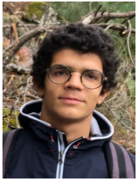
Results

Energy Flow Networks with Particle Distances

Paper in Construction



By: Martim Pinto



Summer Student

MIT-CTP 5689

Moments of Clarity:
Streamlining Latent Spaces in Machine Learning using Moment Pooling

Rikab Gambhir,[1,2,*] Athis Osathapan,[2,3,†] and Jesse Thaler[1,2,‡]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3] *Bowdoin College, Brunswick, ME 04011, USA*

Many machine learning applications involve learning a latent representation of data, which is often high-dimensional and difficult to directly interpret. In this work, we propose "Moment Pooling", a natural extension of Deep Sets networks which drastically decrease latent space dimensionality of these networks while maintaining or even improving performance. Moment Pooling generalizes the summation in Deep Sets to arbitrary multivariate moments, which enables the model to achieve a much higher effective latent dimensionality for a fixed learned latent space dimension. We demonstrate Moment Pooling on the collider physics task of quark/gluon jet classification by extending Energy Flow Networks (EFNs) to Moment EFNs. We find that Moment EFNs with latent dimensions as small as 1 perform similarly to ordinary EFNs with higher latent dimension. This small latent dimension allows for the internal representation to be directly visualized and interpreted, which in turn enables the learned internal jet representation to be extracted in closed form.

[hep-ph] 13 Mar 2024

CONTENTS

I. INTRODUCTION

As modern machine learning (ML) models and their applications continue to grow in size and scope, their internal representations of data become increasingly more complex and difficult to decipher. While there are a variety of ways to interpret what is "learned" in an ML model [1–10], it is often difficult to draw concrete, first-principles conclusions on how these models internally represent learned data, as the latent space tends to be high-dimensional and complex. This, in turn, makes it more

MIT-CTP 5689

Moments of Clarity:
Streamlining Latent Spaces in Machine Learning using Moment Pooling

Rikab Gambhir,[1,2,*] Athis Osathapan,[2,3,†] and Jesse Thaler[1,2,‡]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3] *Bowdoin College, Brunswick, ME 04011, USA*

Many machine learning applications involve learning a latent representation of data, which is often high-dimensional and difficult to directly interpret. In this work, we propose "Moment Pooling", a natural extension of Deep Sets networks which drastically decrease latent space dimensionality of these networks while maintaining or even improving performance. Moment Pooling generalizes the summation in Deep Sets to arbitrary multivariate moments, which enables the model to achieve a much higher effective latent dimensionality for a fixed learned latent space dimension. We demonstrate Moment Pooling on the collider physics task of quark/gluon jet classification by extending Energy Flow Networks (EFNs) to Moment EFNs. We find that Moment EFNs with latent dimensions as small as 1 perform similarly to ordinary EFNs with higher latent dimension. This small latent dimension allows for the internal representation to be directly visualized and interpreted, which in turn enables the learned internal jet representation to be extracted in closed form.

CONTENTS

I. INTRODUCTION

As modern machine learning (ML) models and their applications continue to grow in size and scope, their internal representations of data become increasingly more complex and difficult to decipher. While there are a variety of ways to interpret what is "learned" in an ML model [1–10], it is often difficult to draw concrete, first-principles conclusions on how these models internally represent learned data, as the latent space tends to be high-dimensional and complex. This, in turn, makes it more

$$\mathcal{O}(\mathcal{P}) = F\left(\langle \Phi^a \rangle_{\mathcal{P}}\right)$$

MIT-CTP 5689

Moments of Clarity:
Streamlining Latent Spaces in Machine Learning using Moment Pooling

Rikab Gambhir,[1,2,*] Athis Osathapan,[2,3,†] and Jesse Thaler[1,2,‡]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3] *Bowdoin College, Brunswick, ME 04011, USA*

Many machine learning applications involve learning a latent representation of data, which is often high-dimensional and difficult to directly interpret. In this work, we propose "Moment Pooling", a natural extension of Deep Sets networks which drastically decrease latent space dimensionality of these networks while maintaining or even improving performance. Moment Pooling generalizes the summation in Deep Sets to arbitrary multivariate moments, which enables the model to achieve a much higher effective latent dimensionality for a fixed learned latent space dimension. We demonstrate Moment Pooling on the collider physics task of quark/gluon jet classification by extending Energy Flow Networks (EFNs) to Moment EFNs. We find that Moment EFNs with latent dimensions as small as 1 perform similarly to ordinary EFNs with higher latent dimension. This small latent dimension allows for the internal representation to be directly visualized and interpreted, which in turn enables the learned internal jet representation to be extracted in closed form.

[hep-ph] 13 Mar 2024

CONTENTS

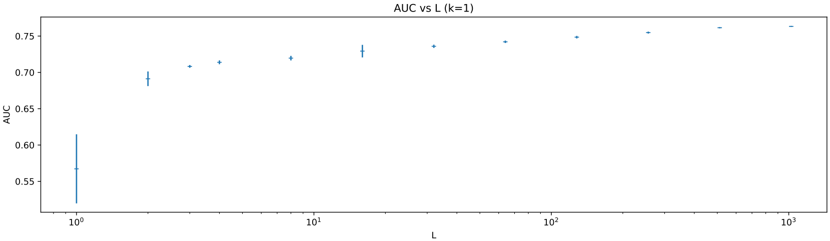
I. INTRODUCTION

As modern machine learning (ML) models and their applications continue to grow in size and scope, their internal representations of data become increasingly more complex and difficult to decipher. While there are a variety of ways to interpret what is "learned" in an ML model [1–10], it is often difficult to draw concrete, first-principles conclusions on how these models internally represent learned data, as the latent space tends to be high-dimensional and complex. This, in turn, makes it more

$$\mathcal{O}\left(\mathcal{P}\right) = F\left(\langle \Phi^a \rangle_{\mathcal{P}}\right) \longrightarrow \mathcal{O}_k(\mathcal{P}) \equiv F\Big(\langle \Phi^a \rangle_{\mathcal{P}}, \langle \Phi^{a_1}\Phi^{a_2} \rangle_{\mathcal{P}}, \dots \langle \Phi^{a_1} \dots \Phi^{a_k} \rangle_{\mathcal{P}} \Big)$$

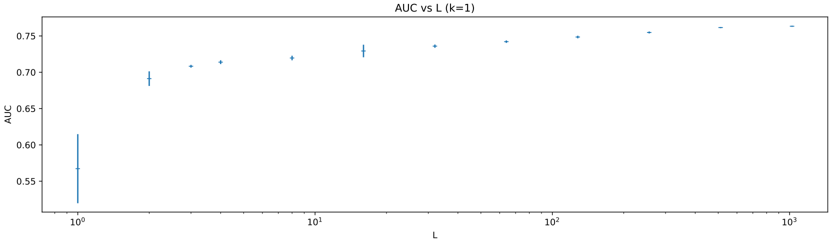
Paper in Construction

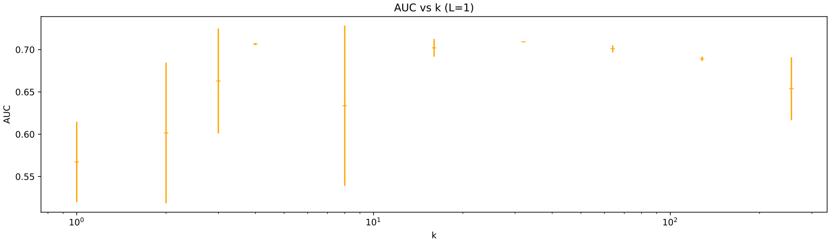
Paper in Construction



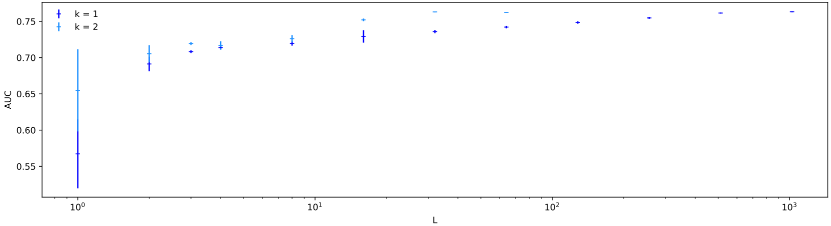
Results
Moments of Clarity

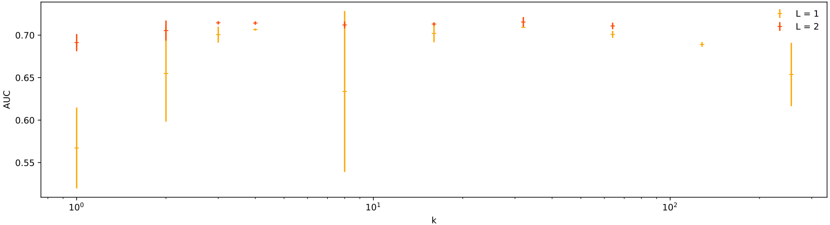
Paper in Construction

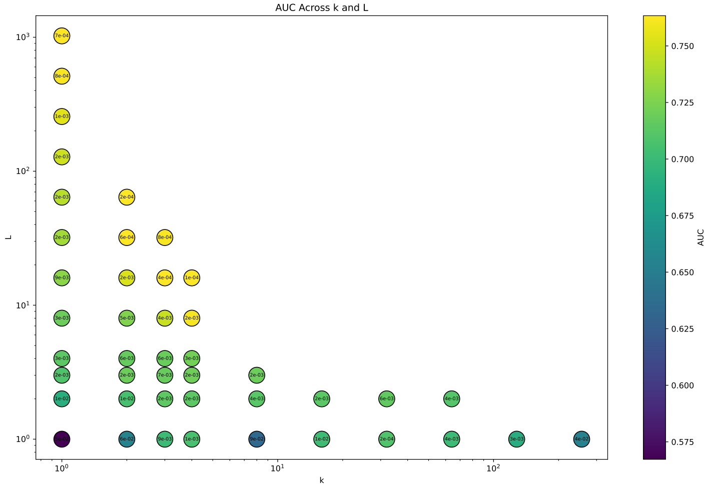




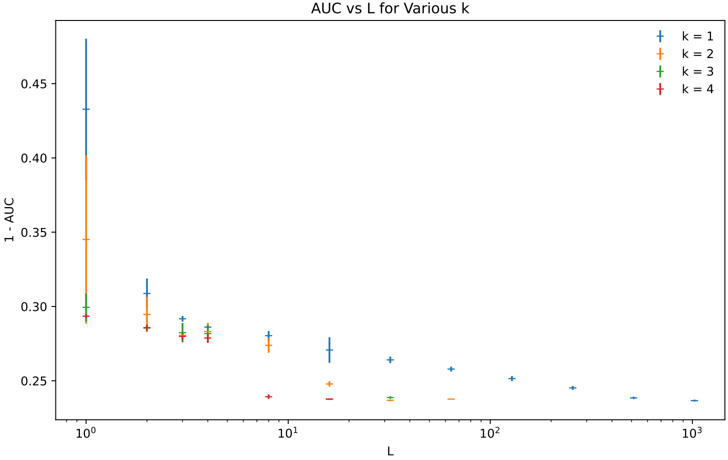
Paper in Construction







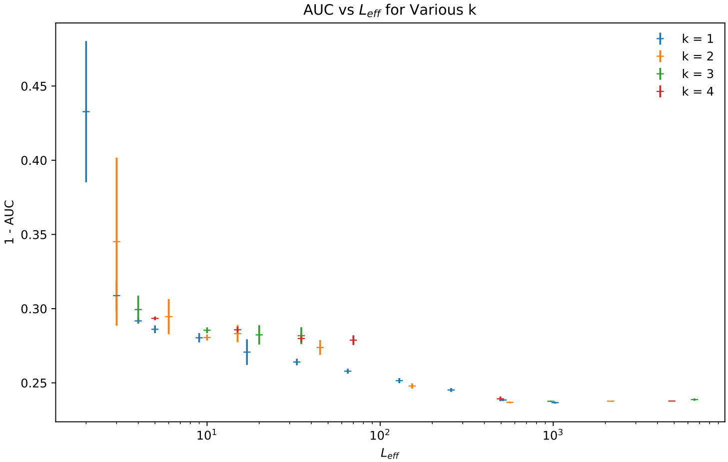
AUC Across k and L
L
k
AUC
0.750
0.725
0.700
0.675
0.650
0.625
0.600
0.575
7e-04
8e-04
1e-03
2e-03
2e-03
2e-04
2e-03
6e-04
8e-04
9e-03
2e-03
4e-04
1e-04
3e-03
5e-03
4e-03
2e-03
3e-03
6e-03
6e-03
3e-03
2e-03
2e-03
7e-03
2e-03
2e-03
1e-02
1e-02
2e-03
2e-03
4e-03
2e-03
6e-03
4e-03
6e-02
9e-03
1e-03
9e-02
1e-02
2e-04
4e-03
3e-03
4e-02



AUC vs L for Various k
k = 1
k = 2
k = 3
k = 4
0.45
0.40
0.35
0.30
0.25
1 - AUC
10^0
10^1
10^2
10^3
L

Results

Moments of Clarity



AUC vs L_{eff} for Various k
k = 1
k = 2
k = 3
k = 4
1 - AUC
0.45
0.40
0.35
0.30
0.25
10^1
10^2
10^3
L_{eff}

Outline

Conclusions and Future Work

Conclusions and Future Work

1. Energy Flow networks and polynomials seem to capture very relevant information for pp versus PbPb discrimination.

Conclusions and Future Work

1. Energy Flow networks and polynomials seem to capture very relevant information for pp versus PbPb discrimination.
2. Adding global jet observables to these complex physics motivated networks seems to be indeed beneficial, significantly the originally improving attained discrimination power.

Conclusions and Future Work

1. Energy Flow networks and polynomials seem to capture very relevant information for pp versus PbPb discrimination.
2. Adding global jet observables to these complex physics motivated networks seems to be indeed beneficial, significantly the originally improving attained discrimination power.
3. Moment Energy Flow Networks open up a new window of exploration of these kinds of models to multiple problems in HEP in general and in jet quenching in particular.

Conclusions and Future Work

1. Energy Flow networks and polynomials seem to capture very relevant information for pp versus PbPb discrimination.
2. Adding global jet observables to these complex physics motivated networks seems to be indeed beneficial, significantly the originally improving attained discrimination power.
3. Moment Energy Flow Networks open up a new window of exploration of these kinds of models to multiple problems in HEP in general and in jet quenching in particular.
4. Future work will focus on adding observables to Moment EFNs and attempting to obtain an interpretable latent space and perhaps relating it to calculable observables well under theoretical control.

Conclusions and Future Work

1. Energy Flow networks and polynomials seem to capture very relevant information for pp versus PbPb discrimination.
2. Adding global jet observables to these complex physics motivated networks seems to be indeed beneficial, significantly the originally improving attained discrimination power.
3. Moment Energy Flow Networks open up a new window of exploration of these kinds of models to multiple problems in HEP in general and in jet quenching in particular.
4. Future work will focus on adding observables to Moment EFNs and attempting to obtain an interpretable latent space and perhaps relating it to calculable observables well under theoretical control.
5. What observable will the network learn if we give it all the observables we know are useful? Can we calculate it from first principles? Can we relate this to the quenching phenomena?

Thank you for your attention!

Questions?

Questions?

UE

Gen and Rec
UE Gen
UE Fits
UE Comp
Sub Dets
Sub Qual
US Steps
US Plots
UE Obs
UE ML

Energy Flow

Obs.
NSub LDA
NSub DNN
EFP LDA
EFP DNN
EFNs
AUC Error
Weights
Model comp
Preproc
Model archs
Obs. Form.

Backup

Underlying Event Contamination

Generation Details		
Process		dijets
Centrality		$[0, 10]\%$
τ_i	$=$	0.4
T_i	$=$	590 MeV
$\sqrt{s_{NN}}$	$=$	5.02 TeV
$\hat{p}_t$	$>$	50 GeV
$\lvert\eta\rvert$	$<$	4

Reconstruction Details		
p_t^{part}	$>$	100 MeV
$\lvert\eta^{part}\rvert$	$<$	4
Jets		0.4 anti_kt
$\lvert\eta^{jets}\rvert$	$<$	3
$\Delta\phi$	$<$	$5\pi/6$
p_t^{lead}	$>$	120 GeV
$p_t^{sublead}$	$>$	50 GeV

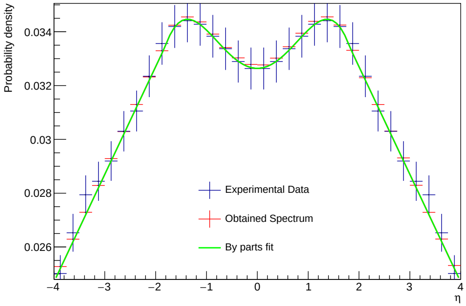
Experimentally motivated UE generation steps:

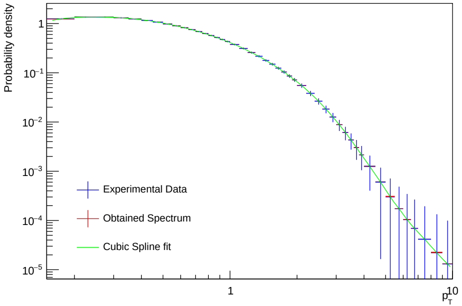
1. Fit the pseudo-rapidity distribution of the UE measured experimentally from [1]. We have used a polynomial fit.
2. Fit the transverse momenta distribution of the UE measured experimentally in [2]. We have used a cubic spline.
3. Take the ϕ distribution to be uniform.
4. Take the number of particles per UE to follow a Gaussian distribution of experimentally motivated average value and standard deviation.
5. For each particle to be generated, sample a value for p_T, η and ϕ from the considered distributions.
6. Considering only pions, sample randomly and uniformly one of the three species, and use its mass to complete the four-momentum of the particle.

[1] Phys.Lett.B 772 (2017) 567-577, 2017.
[2] JHEP 11 (2018) 013, 2018.

UE fits

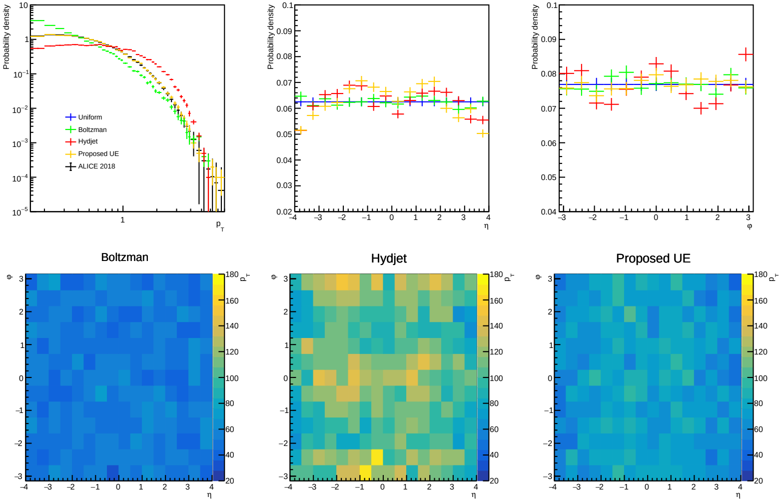
Internal Revision



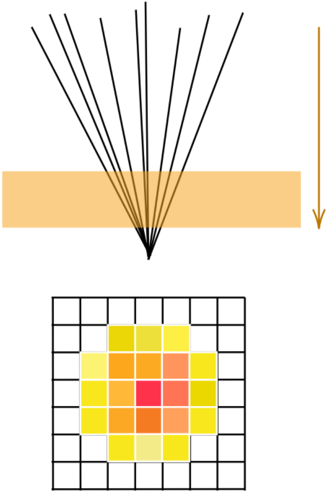


UE Comparison

Internal Revision



Subtraction Details



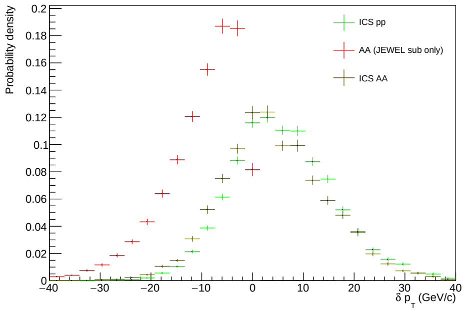
We have performed two different types of subtractions:

1. JEWEL's internal background subtraction to give physical medium response (only for PbPb and this is always performed before embedding) [3]
2. Iterative Constituent Subtraction of UE which we apply to both pp and PbPb embedded events [4]

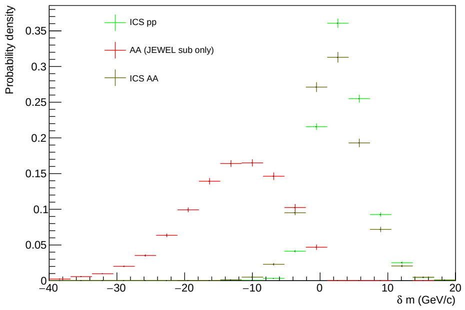
We have used the parameters suggested in [4] for 0.4 anti-kt jets.

[3] Eur.Phys.J.C 82 (2022) 11, 1010
[4] JHEP 08 (2019) 175

Subtraction quality plots



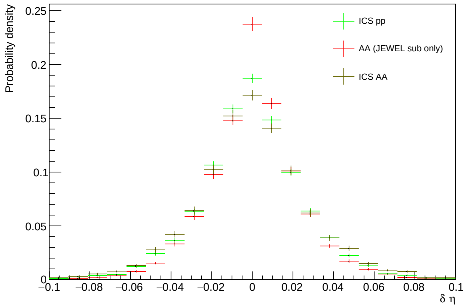
$$\delta p_T = p_T^{sub} - p_T^{gen}$$



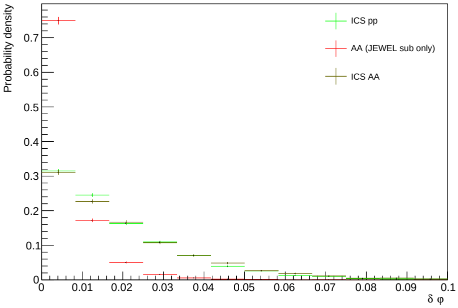
$$\delta m = m^{sub} - m^{gen}$$

Subtraction quality plots

Internal Revision



$$\delta\eta = \eta^{sub} - \eta^{gen}$$



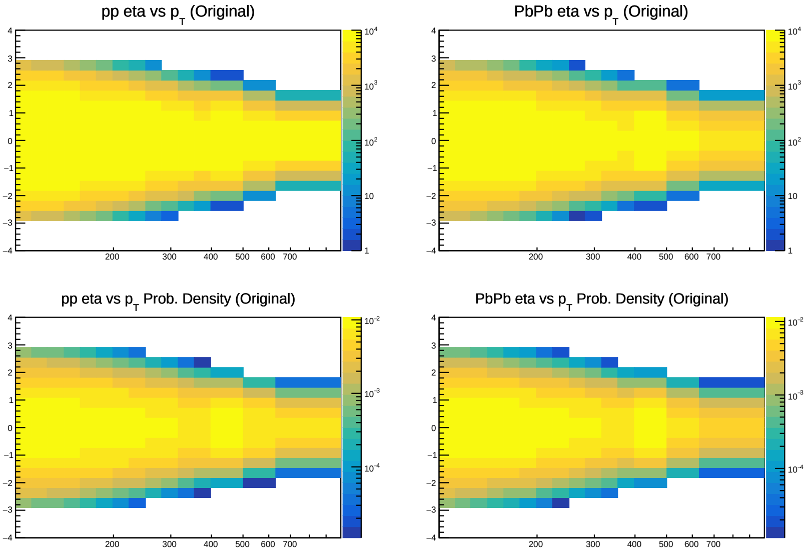
$$\delta\phi = \phi^{sub} - \phi^{gen}$$

Undersampling steps:

1. Bin pp and PbPb data in p_T and η.
2. Check if there are bins with more PbPb events than pp.
3. If so remove randomly and uniformly events from PbPb until no bin has a larger population of PbPb compared to pp.
4. For all bins, remove randomly pp events from each until the number of pp events matches the number of PbPb events in each and every bin.

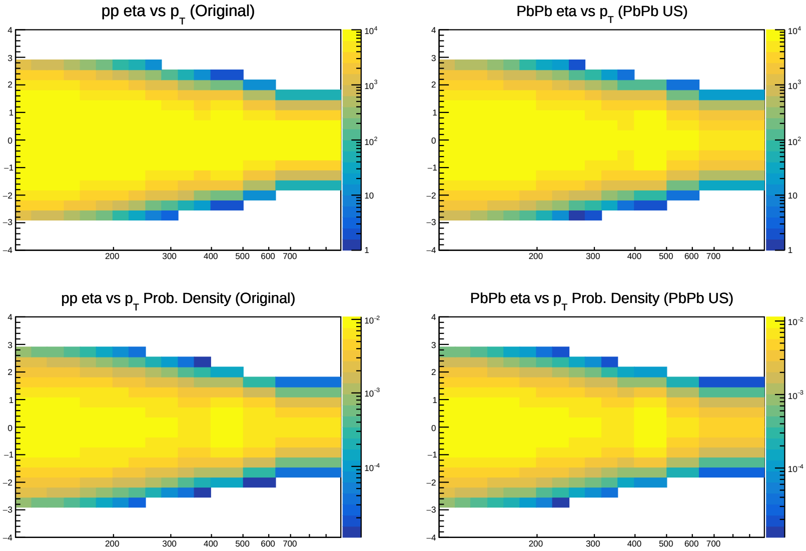
Undersampling Original plot

Internal Revision



Undersampling PbPb step plot

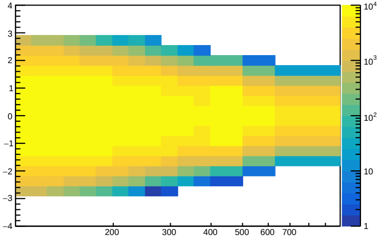
Internal Revision



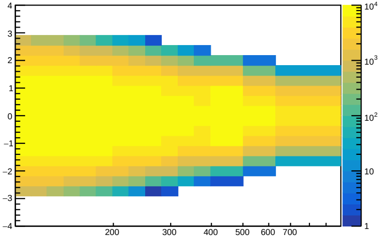
Undersampling final plot

Internal Revision

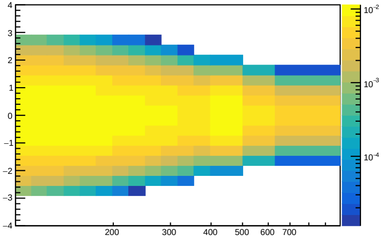
pp eta vs p_T (pp US)



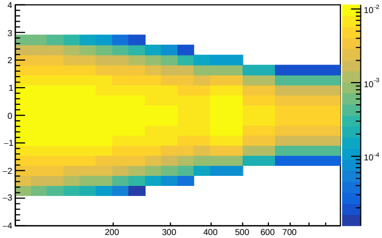
PbPb eta vs p_T (PbPb US)



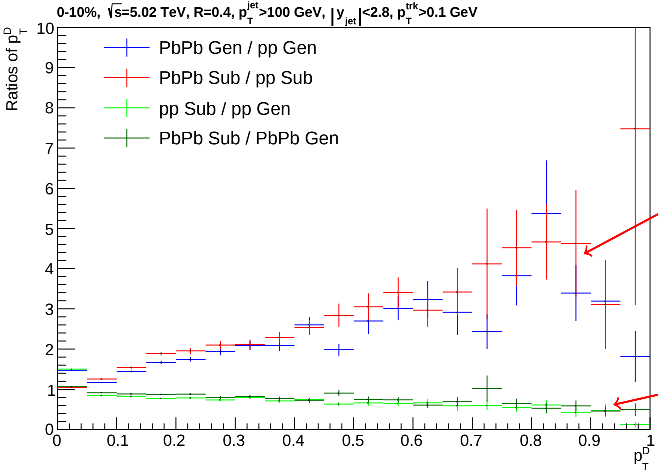
pp eta vs p_T Prob. Density (pp US)



PbPb eta vs p_T Prob. Density (PbPb US)



UE contamination effect on p_T^D



Robust to UE contamination

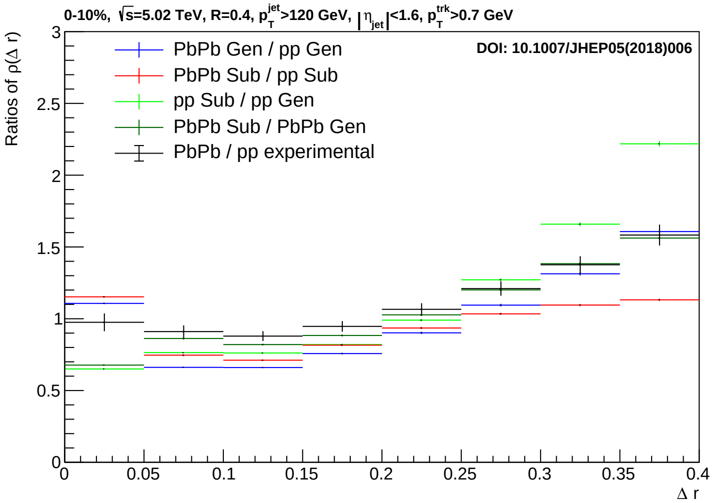
UE contamination effects for pp and PbPb

$$p_T^D = \left(\frac{p_T^i}{p_T^{jet}}\right)^2 cos^2(r);\, r = \sqrt{(\phi_i - \phi_{jet})^2 + (\eta_i - \eta_{jet})^2}$$

Internal Revision

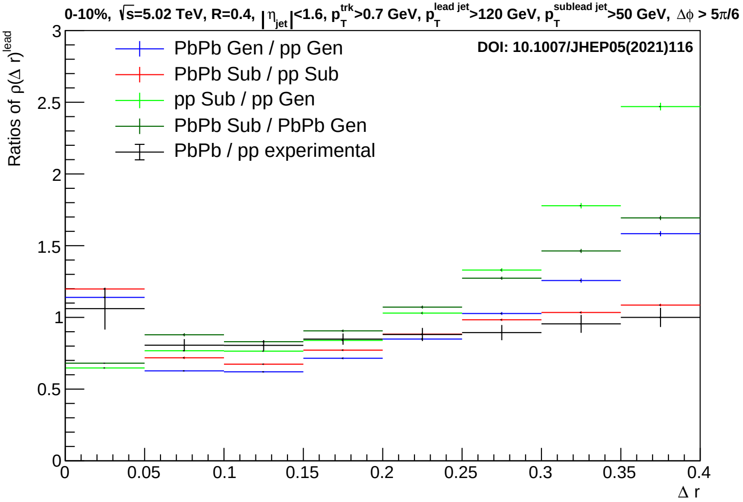
UE contamination effect on inclusive jet profile (ρ)

Internal Revision



$$\rho(\Delta r) = \frac{1}{\delta r} \frac{1}{N_{jets}} \sum_{consts \in \Delta r} p_T^{const}$$

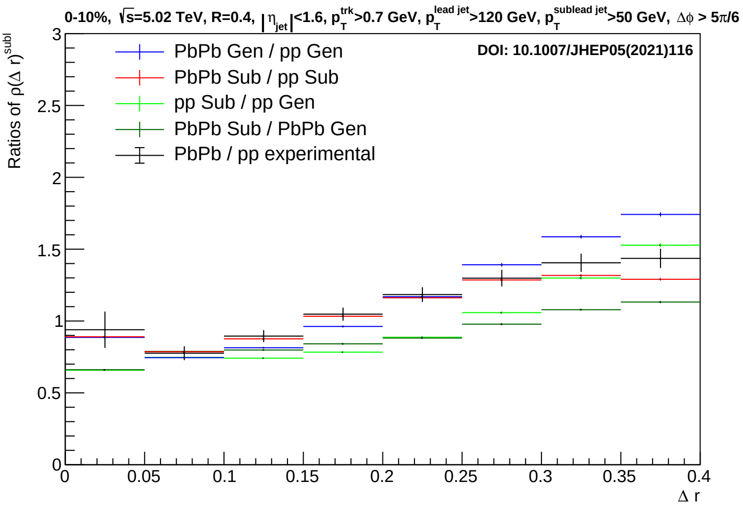
UE contamination effect on jet profile of leading jet (ρ^{lead})



Internal Revision

$$\rho(\Delta r) = \frac{1}{\delta r}\frac{1}{N_{jets}} \sum_{consts \in \Delta r} p_T^{const}$$

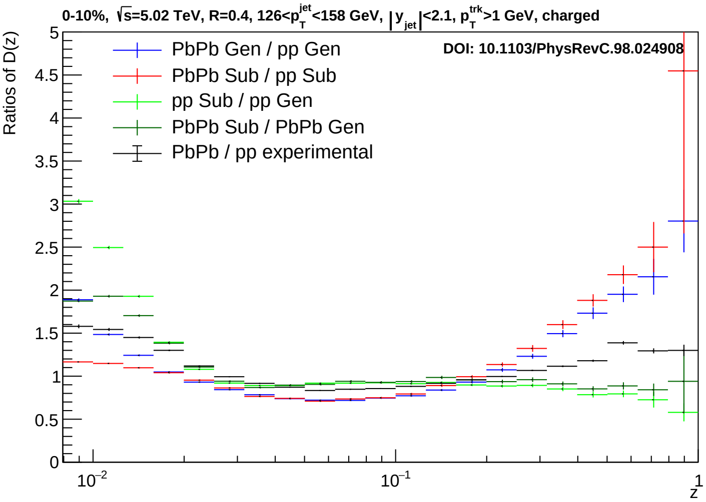
UE contamination effect on jet profile of subleading jet (ρ^{subl})



Internal Revision

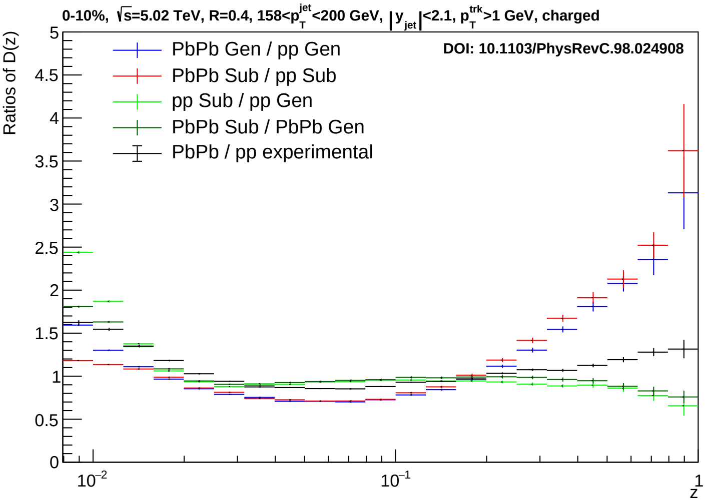
$$\rho(\Delta r) = \frac{1}{\delta r}\frac{1}{N_{jets}}\sum_{consts \in \Delta r} p_T^{const}$$

UE contamination effect on jet fragmentation functions ($D(z)$)



$$D(z) = \frac{1}{N_{jet}} \frac{dN_{chg}}{dz}; z = \frac{p_T^{const} cos(\Delta R)}{p_T^{jet}}$$

UE contamination effect on jet fragmentation functions ($D(z)$)

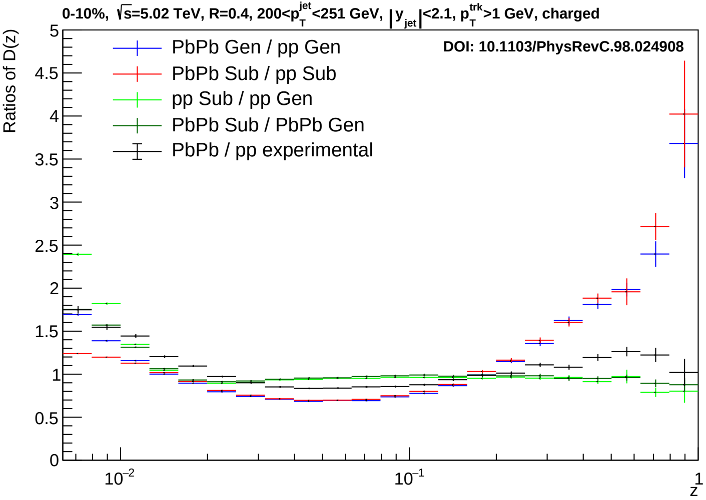


$$D(z) = \frac{1}{N_{jet}}\frac{dN_{chg}}{dz};\ z = \frac{p_T^{const}\cos(\Delta R)}{p_T^{jet}}$$

Internal Revision

UE contamination effect on jet fragmentation functions ($D(z)$)

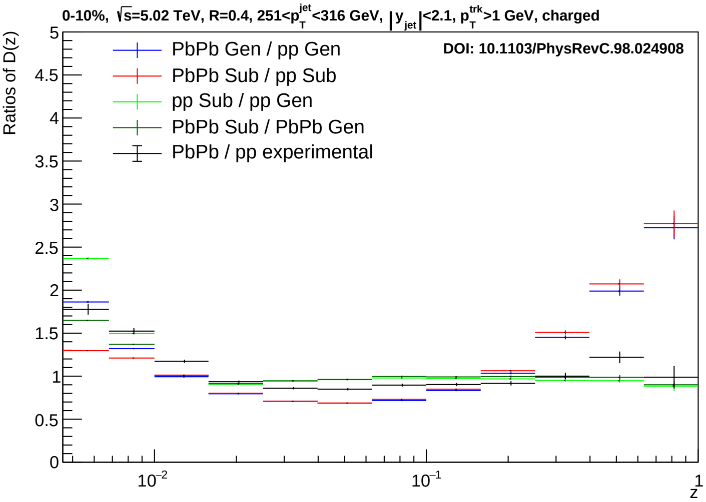
Internal Revision



$$D(z) = \frac{1}{N_{jet}} \frac{dN_{chg}}{dz}; \; z = \frac{p_T^{const} cos(\Delta R)}{p_T^{jet}}$$

UE contamination effect on jet fragmentation functions ($D(z)$)

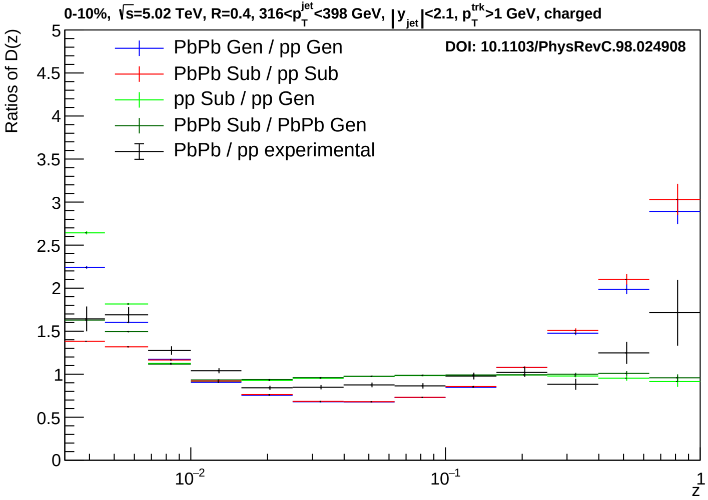
Internal Revision



$$D(z) = \frac{1}{N_{jet}} \frac{dN_{chg}}{dz}; z = \frac{p_T^{const} \cos(\Delta R)}{p_T^{jet}}$$

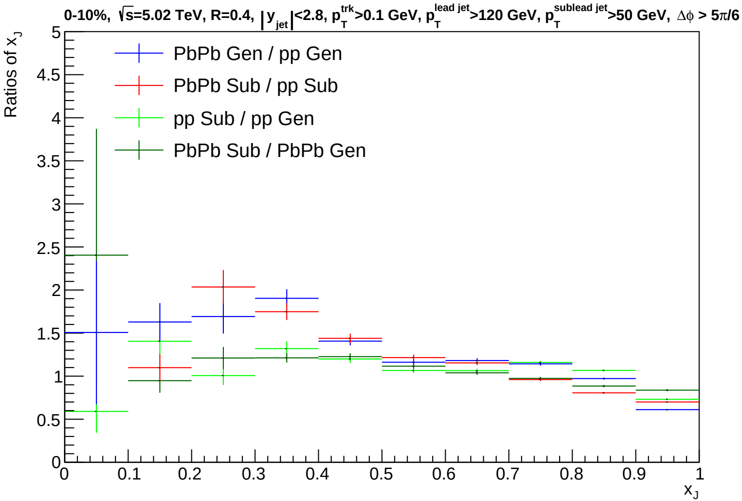
UE contamination effect on jet fragmentation functions ($D(z)$)

Internal Revision



$$D(z) = \frac{1}{N_{jet}}\frac{dN_{chg}}{dz};\, z = \frac{p_T^{const}\cos(\Delta R)}{p_T^{jet}}$$

UE contamination effect on the dijet asymmetry (x_J)

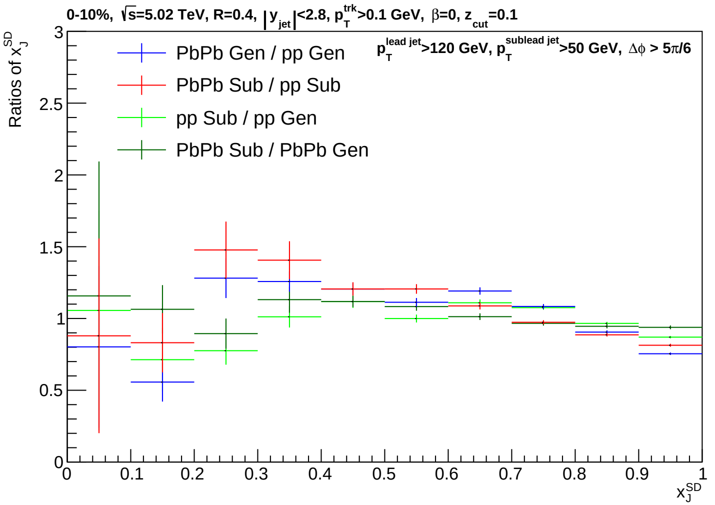


$$x_j = p_T^{sublead} / p_T^{lead}$$

Internal Revision

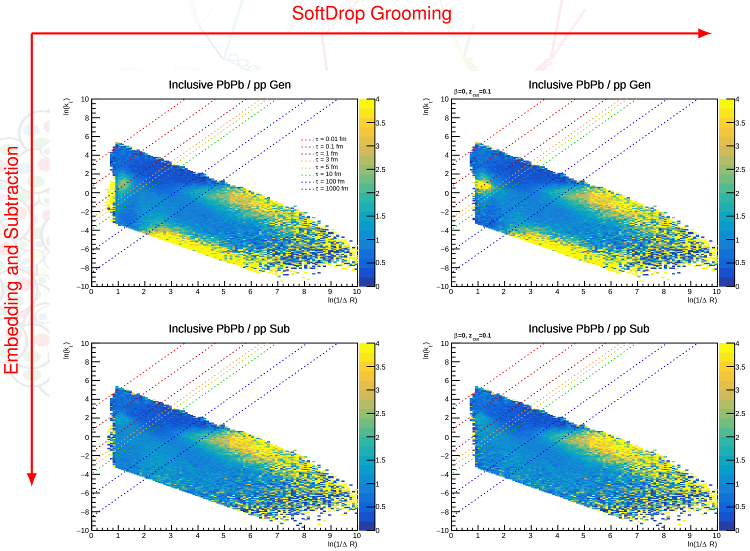
UE contamination effect on the groomed dijet asymmetry (x_J^{SD})

Internal Revision



$$x_j = p_T^{sublead} / p_T^{lead}$$

UE contamination effect on lund planes



Internal Revision

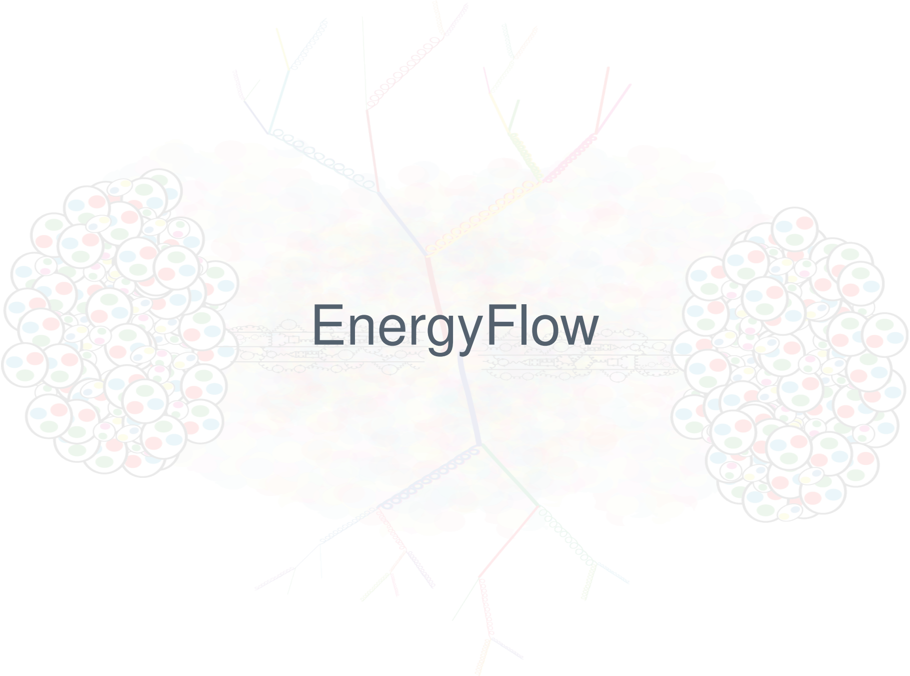
Grooming seems to increase the signal in the medium time window, but the subtraction always depletes the signal in this region.

Observables used for the inital study of the UE contamination in ML

Observable	Type
y_{SD} ϕ_{SD} $\Delta p_{T,SD} = p_{T,jet} - p_{T,jet_{SD}}$ m_{SD} $n_{\text{const,SD}}$	Jet Momenta and Constituent Multiplicity
$\bar{r}_{SD} = \frac{1}{n_{\text{const,SD}}}\lambda^0_{1,SD}$ $\bar{r}^2_{SD} = \frac{1}{n_{\text{const,SD}}}\lambda^0_{2,SD}$ $rz_{SD} = \lambda^1_{1,SD}$ $r^2z_{SD} = \lambda^1_{2,SD}$ $\bar{z}^2_{SD} = \frac{1}{n_{\text{const,SD}}}\lambda^2_{0,SD}$ $p_TD_{SD} = \sqrt{\sum_{i\in jet_{SD}} p^2_{T,i}}/p_{T,jet,SD}$	Angularities
$\tau_{2,SD},\ \tau_{3,SD}$ $\tau_{1,2,SD},\ \tau_{2,3,SD}$	N-subjettiness
$\lvert Q^{0.3}_{SD}\rvert,\ \lvert Q^{0.5}_{SD}\rvert,\ \lvert Q^{0.7}_{SD}\rvert,\ \lvert Q^{1.0}_{SD}\rvert,$	Jet-Charges
$R_g,\ z_g,\ n_{SD}$	SoftDrop Grooming Intrinsic
$R_{g,A},\ z_{g,A},\ \kappa_A$ with $A \in \{TD, ktD, zD\}$	Dynamical Grooming Intrinsic

[5] 10.48550/arXiv.2304.07196

EnergyFlow



EFP selected quotes from the paper

"These observables are multiparticle energy correlators with specific angular structures which directly result from IRC safety."

"EFPs can be viewed as a discrete set of C-correlators"

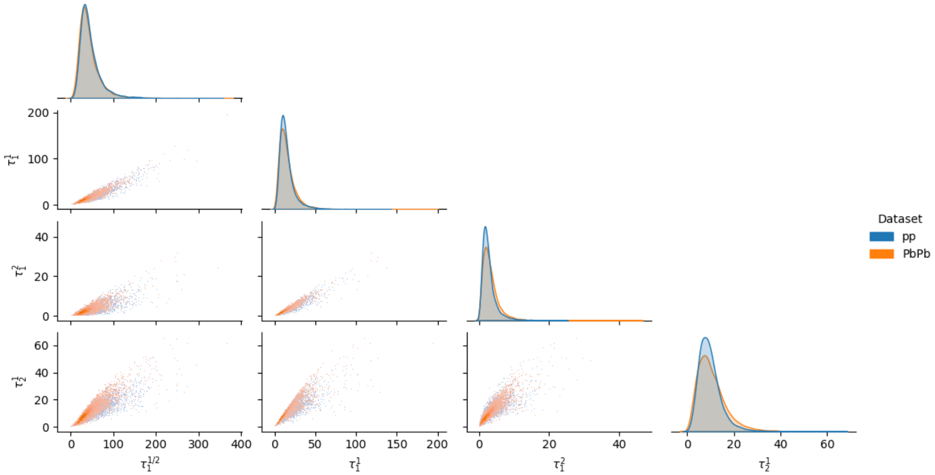
"EFPs form a linear basis of all IRC-safe observables, making them suitable for a wide variety of jet substructure contexts where linear methods are applicable"

"There is a one-to-one correspondence between EFPs and loopless multigraphs, which helps to visualize and calculate the EFPs"

"(...) we usually truncate by restricting to the set of all multigraphs with at most d edges (...) this truncation results in a finite number of EFPs at each order of truncation, which is not true for truncation by the number of vertices."

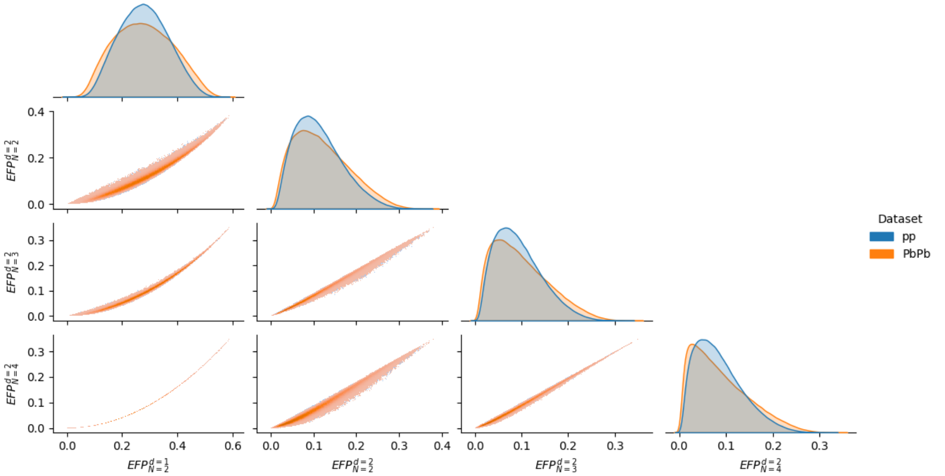
[6] doi.org/10.1007/JHEP04(2018)013

N-Subjetiness Pairplot



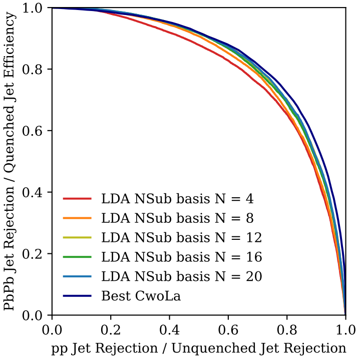
EFP Pairplot

Paper in Construction

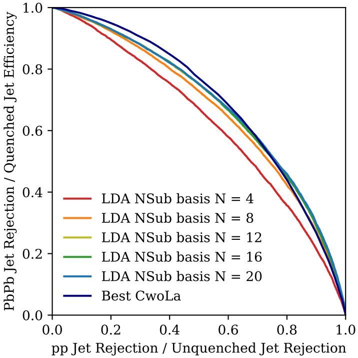


N-Subjetiness LDA ROCs

No UE



With subtracted UE

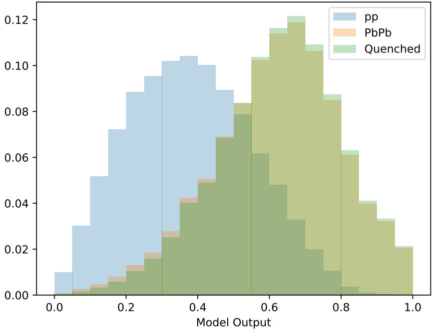


[6] doi.org/10.1007/JHEP04(2018)013

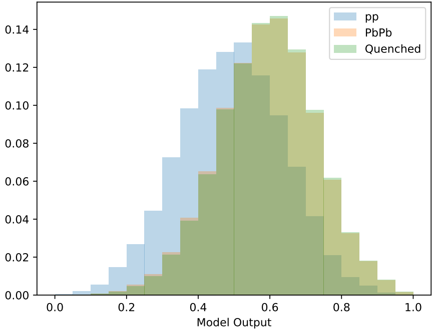
N-Subjetiness LDA Model Output

Paper in Construction

No UE



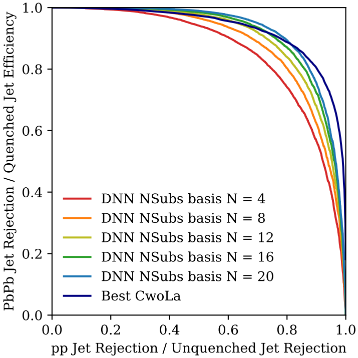
With subtracted UE



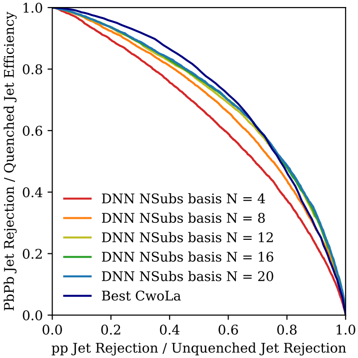
N-Subjetiness DNN ROCs

Paper in Construction

No UE



With subtracted UE

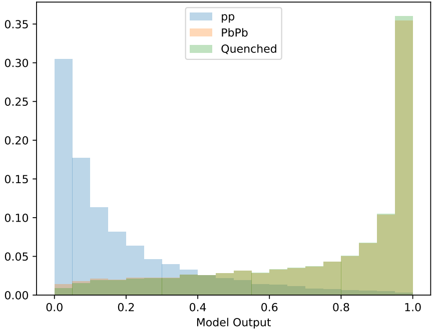


[6] doi.org/10.1007/JHEP04(2018)013

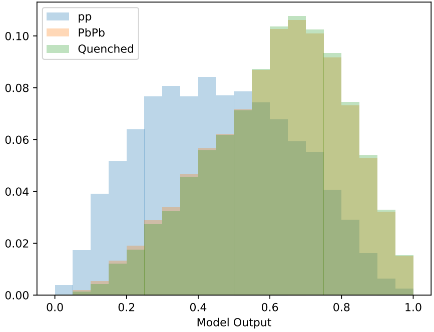
N-Subjetiness DNN Model Output

Paper in Construction

No UE



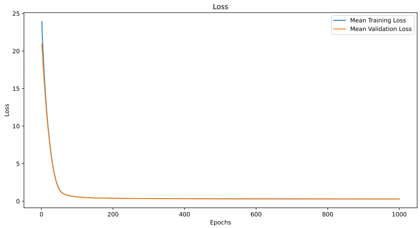
With subtracted UE



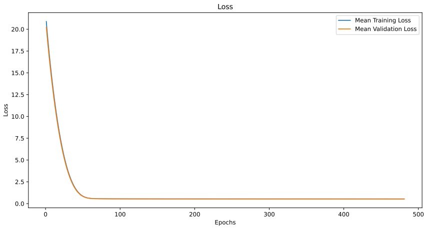
N-Subjetiness DNN Loss

Paper in Construction

No UE



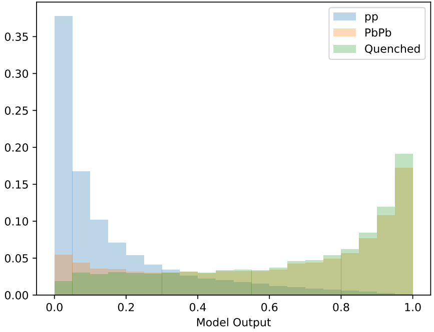
With subtracted UE



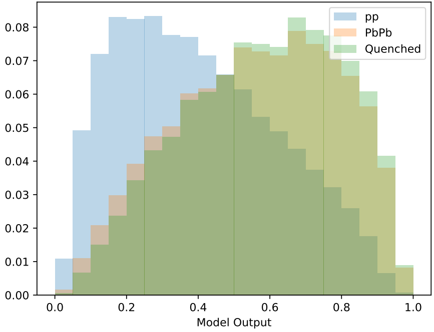
EFP LDA Model Output

Paper in Construction

No UE



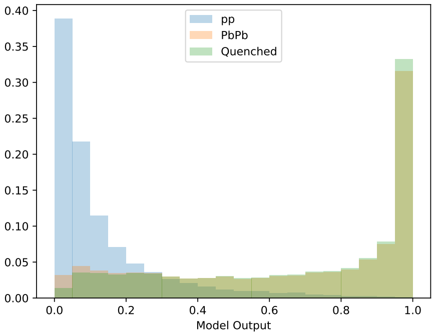
With subtracted UE



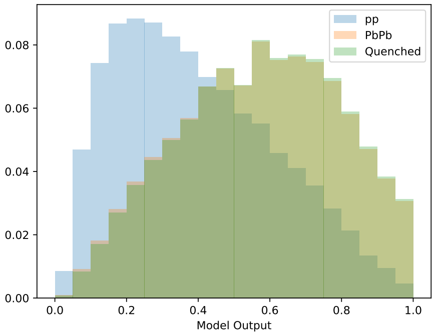
EFP DNN Model Output

Paper in Construction

No UE



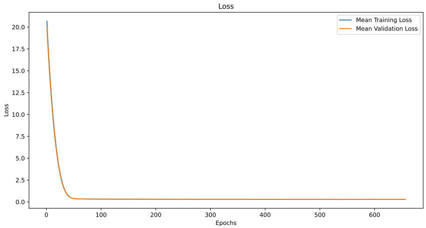
With subtracted UE



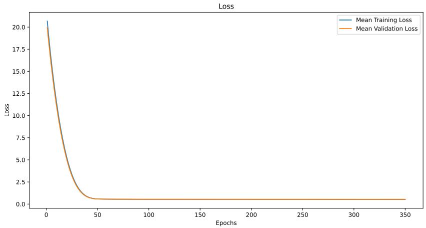
EFP DNN Loss

Paper in Construction

No UE



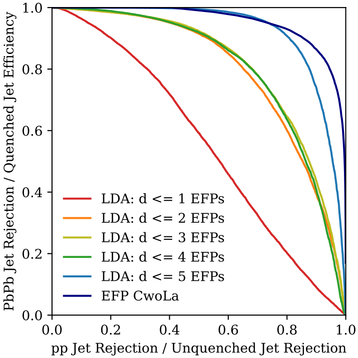
With subtracted UE



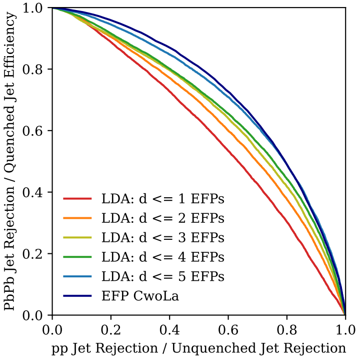
EFP Extended LDA ROCs

Paper in Construction

No UE



With subtracted UE

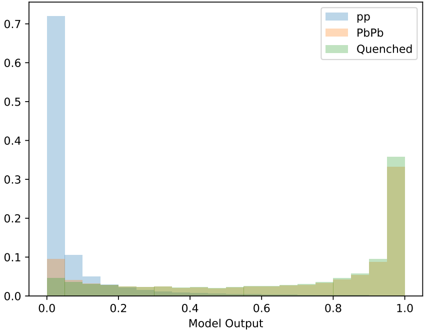


[6] doi.org/10.1007/JHEP04(2018)013

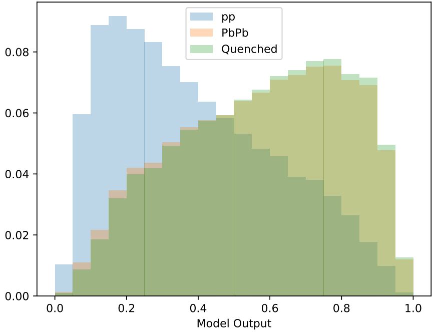
EFP Extended LDA Model Output

Paper in Construction

No UE



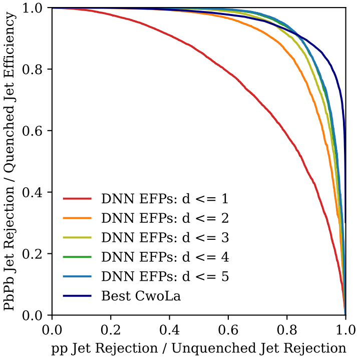
With subtracted UE



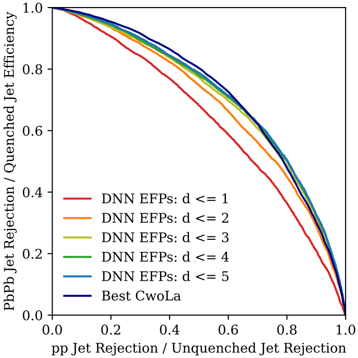
EFP Extended DNN ROCs

Paper in Construction

No UE



With subtracted UE

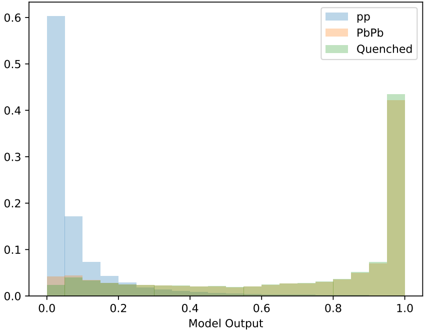


[6] doi.org/10.1007/JHEP04(2018)013

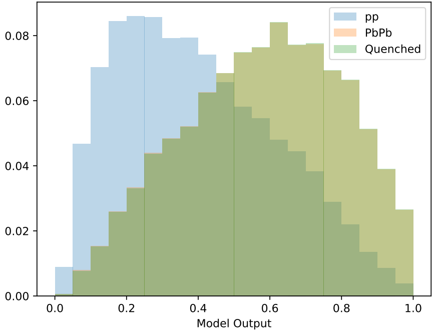
EFP Extended DNN Model Output

Paper in Construction

No UE



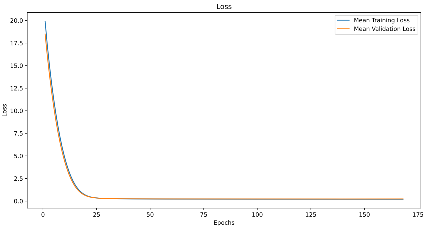
With subtracted UE



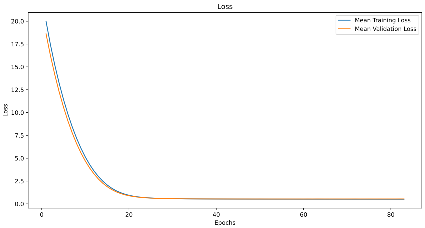
EFP Extended DNN Loss

Paper in Construction

No UE

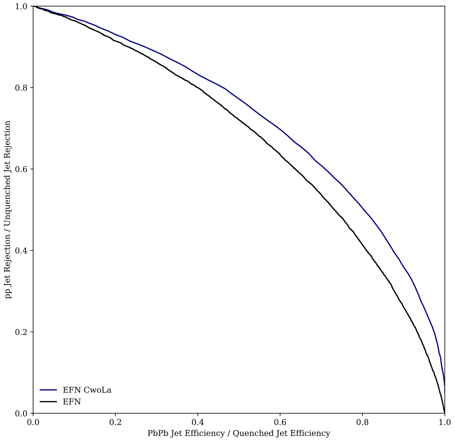


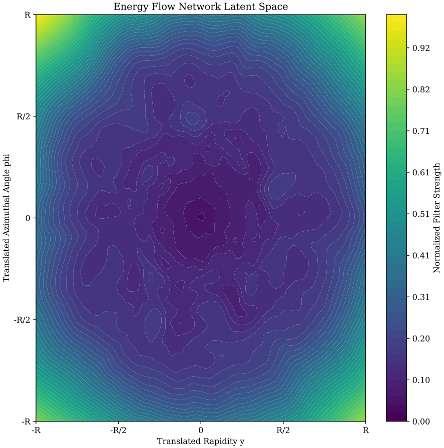
With subtracted UE



EFN ROC and Latent Space

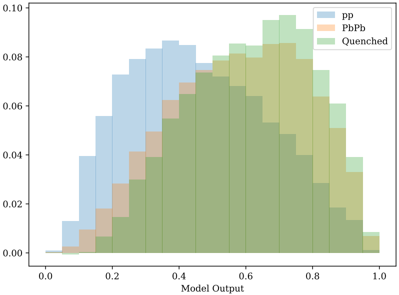
Paper in Construction

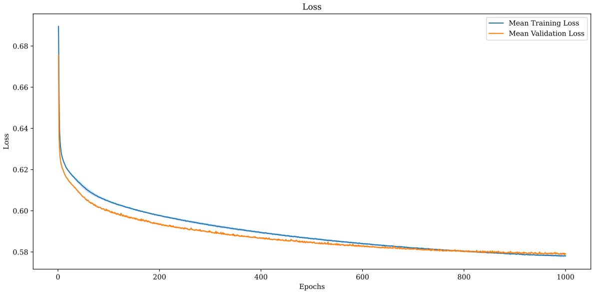




EFN Model Output and Loss

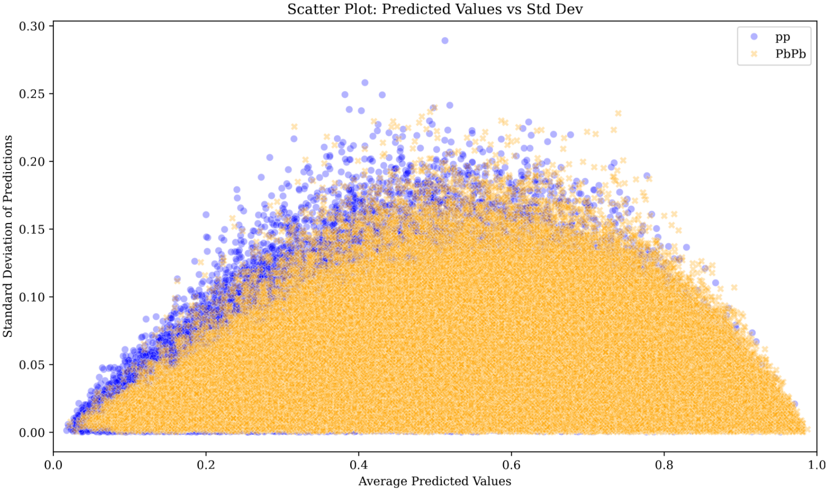
Paper in Construction





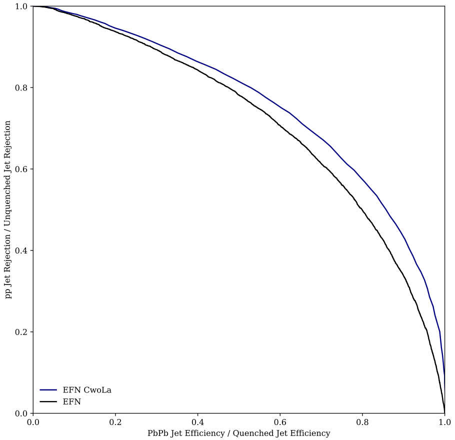
EFN Output Mean vs Std. Dev across folds

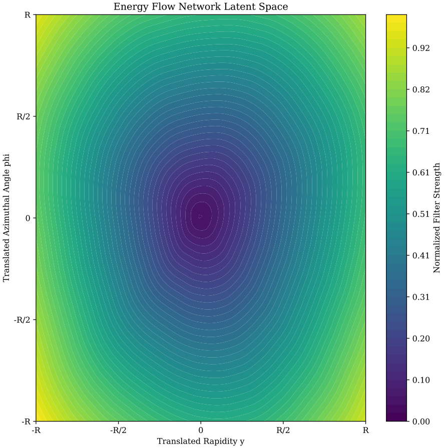
Paper in Construction



EFN + EFP ROC and Latent Space

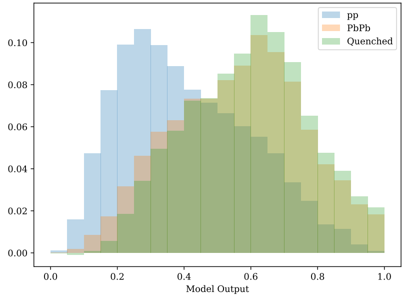
Paper in Construction

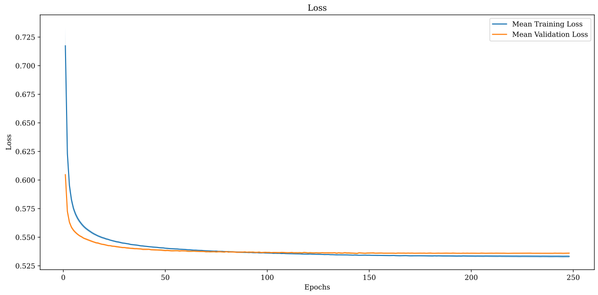




EFN + EFP Model Output and Loss

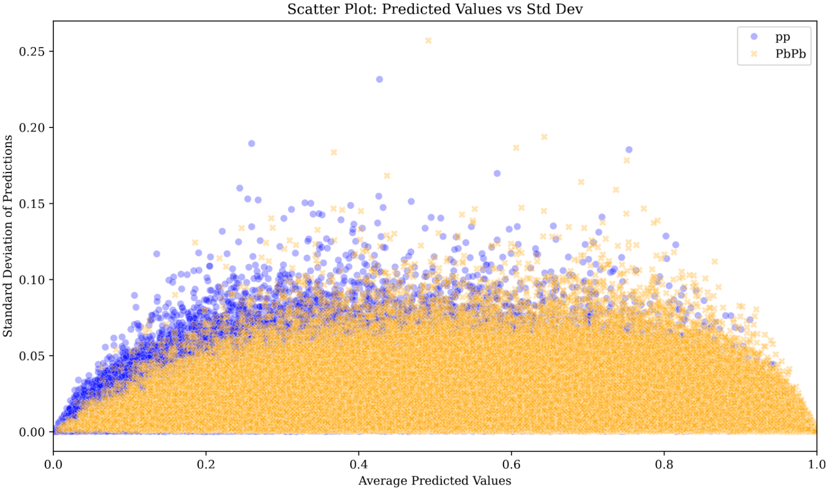
Paper in Construction





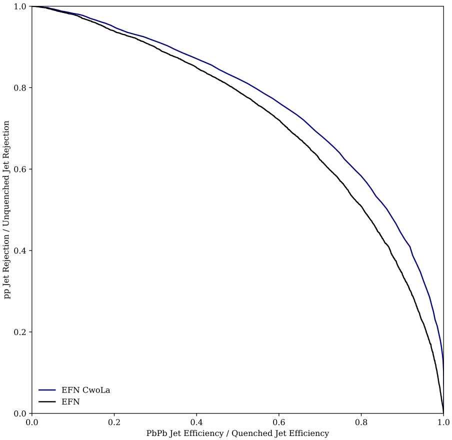
EFN + EFP Output Mean vs Std. Dev Across Folds

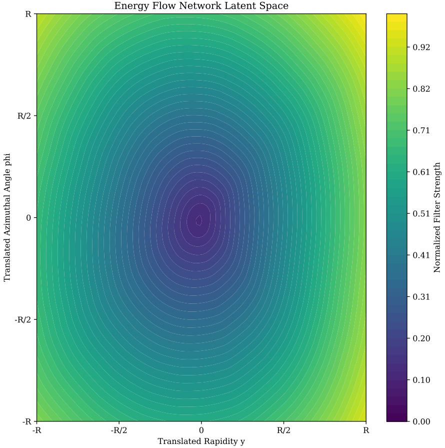
Paper in Construction



EFN + EFP Ext. ROC and Latent Space

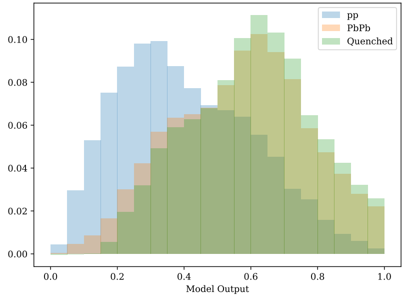
Paper in Construction

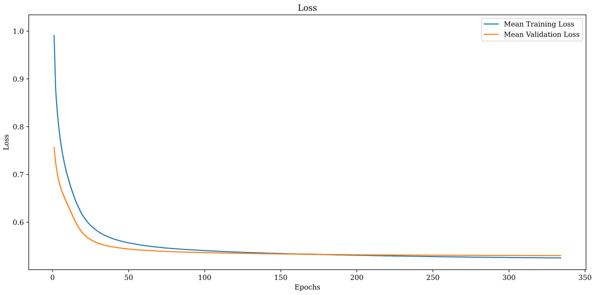




EFN + EFP Ext. Model Output and Loss

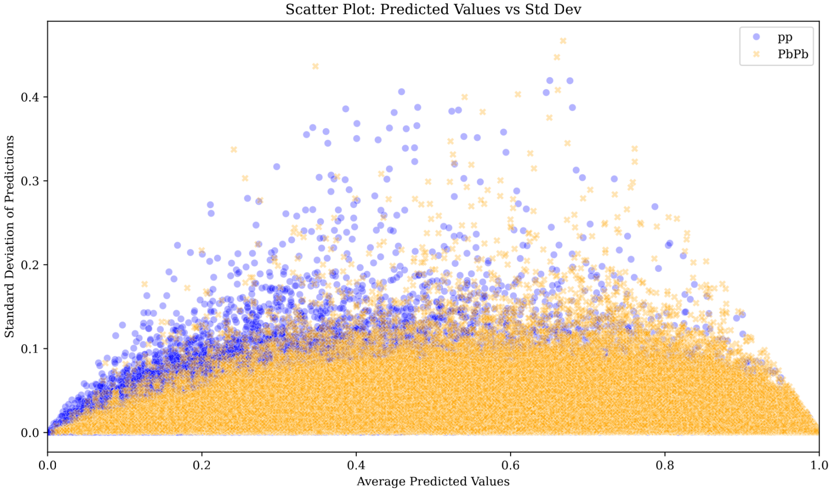
Paper in Construction





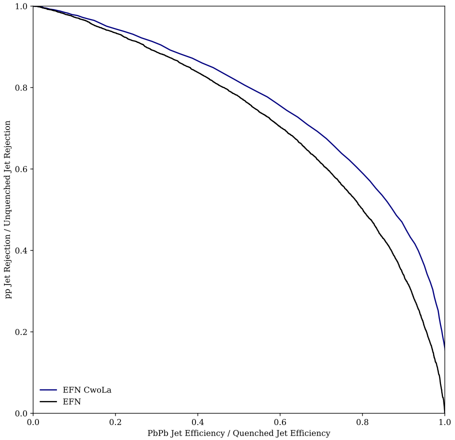
EFN + EFP Ext. Output Mean vs Std. Dev

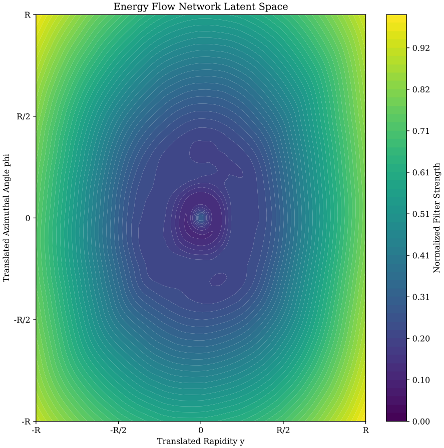
Paper in Construction



EFN + NSubs ROC and Latent Space

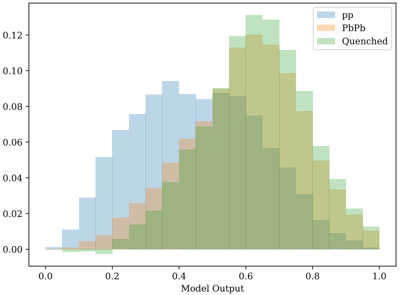
Paper in Construction

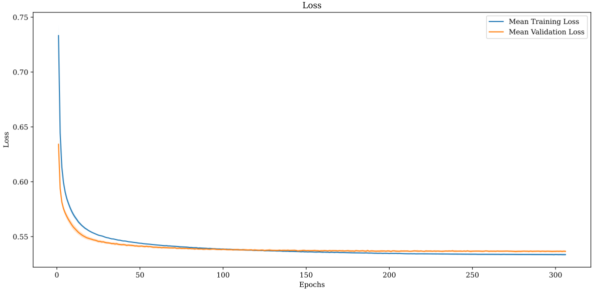




EFN + NSub Model Output and Loss

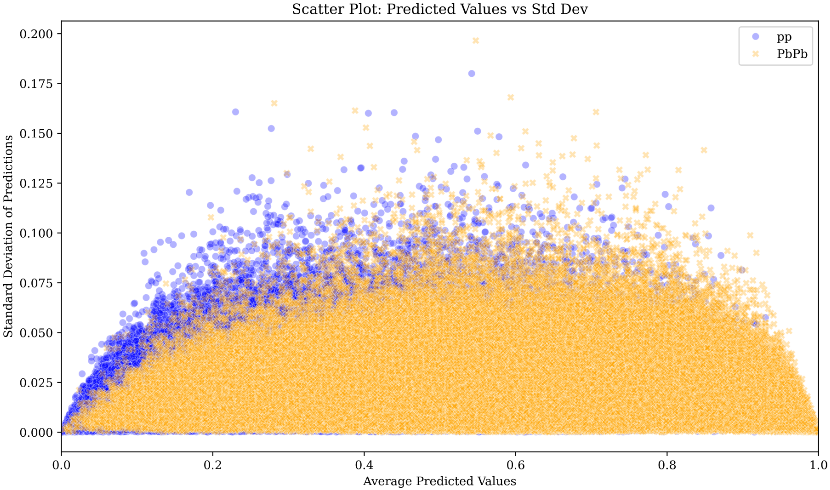
Paper in Construction





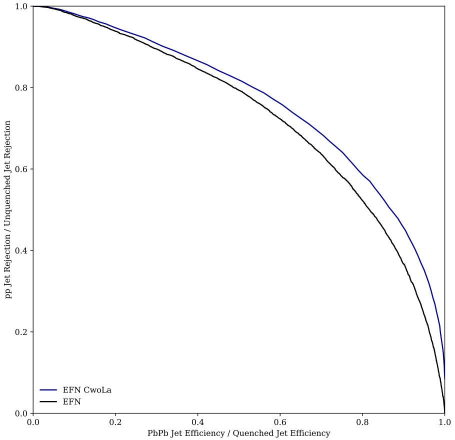
EFN + NSub Output Mean vs Std. Dev

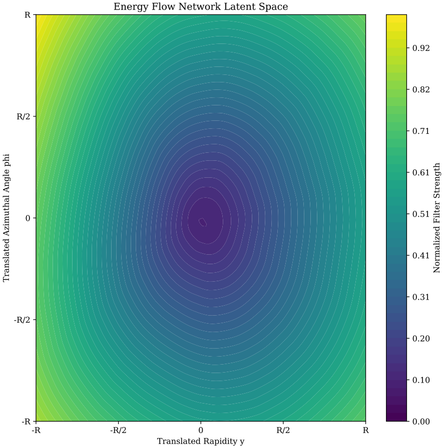
Paper in Construction



EFN + EFP + NSub ROC and Latent Space

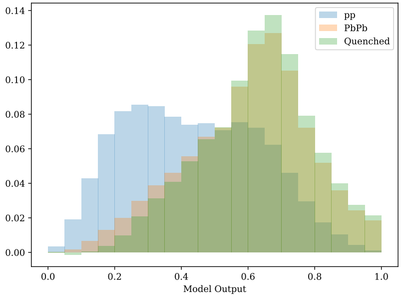
Paper in Construction

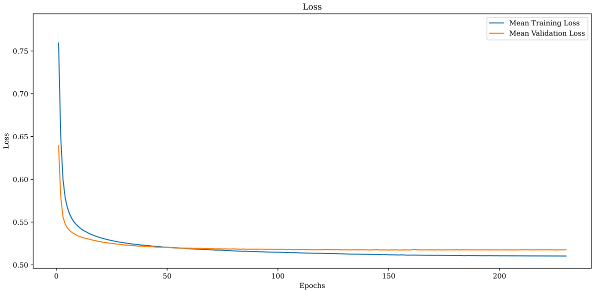




EFN + EFP + NSub Model Output and Loss

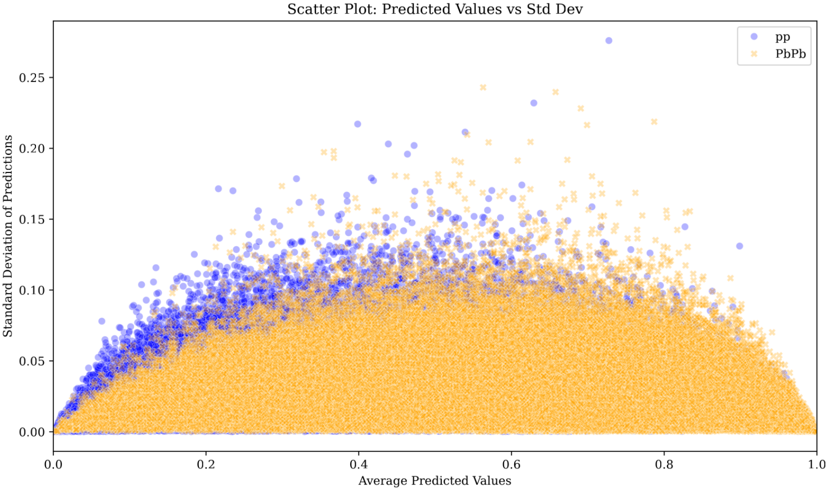
Paper in Construction





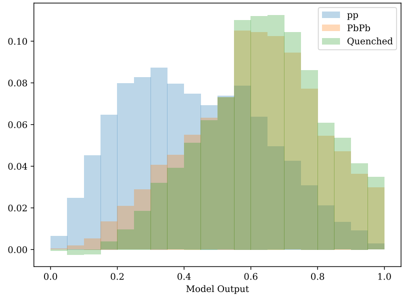
EFN + EFP + NSub Output Mean vs Std. Dev

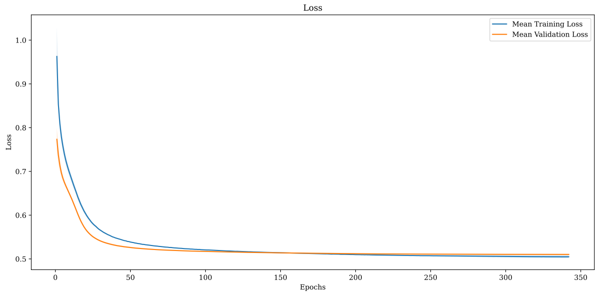
Paper in Construction



EFN + EFP Ext. + NSub Model Output and Loss

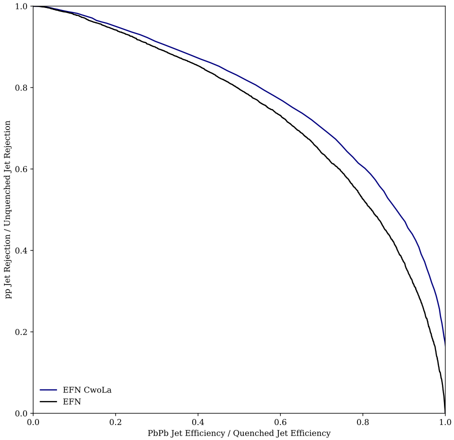
Paper in Construction

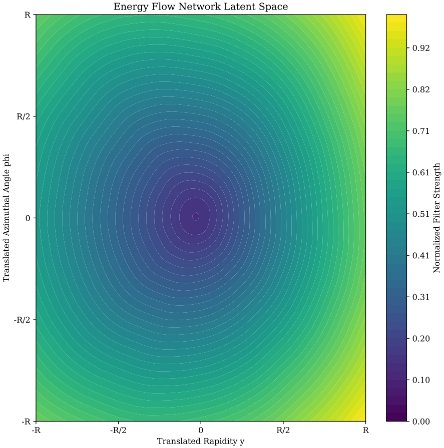




EFN + EFP Ext. + NSubs ROC and Latent Space

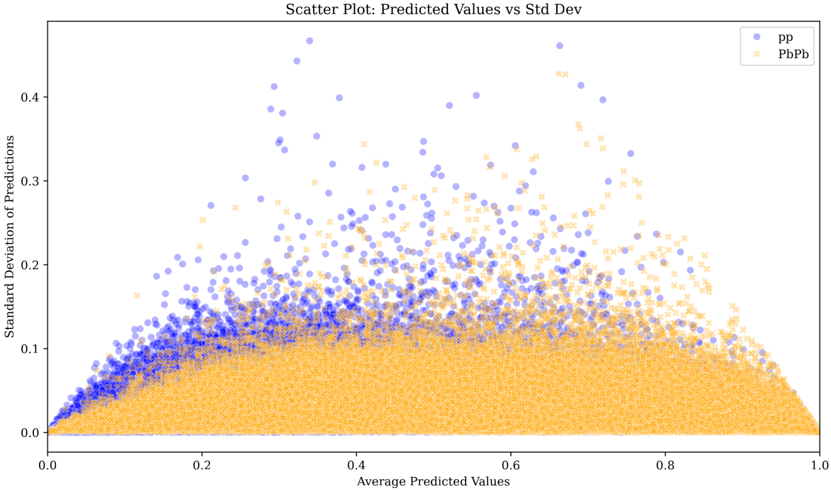
Paper in Construction





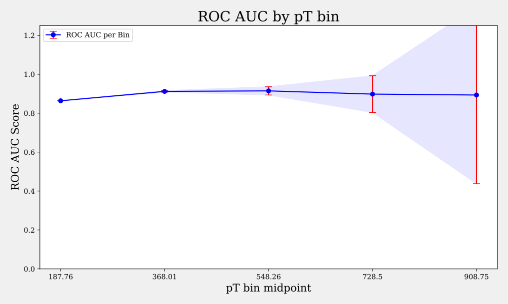
EFN + EFP Ext. + NSub Output Mean vs Std. Dev

Paper in Construction

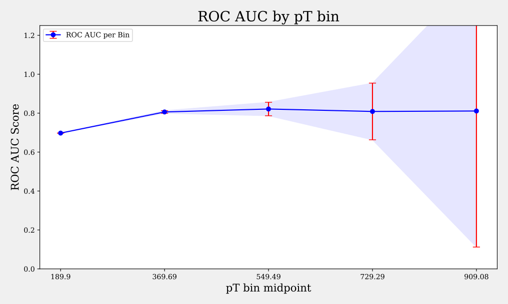


Thinking about the ROC AUC Error

No UE



With subtracted UE



Errors given by:

$$Q_1 = \frac{AUC}{2 - AUC}, \quad Q_2 = \frac{2 \cdot AUC^2}{1 + AUC}$$

$$SE_{AUC} = \sqrt{\frac{AUC \cdot (1 - AUC) + (n_{pos} - 1) \cdot (Q_1 - AUC^2) + (n_{neg} - 1) \cdot (Q_2 - AUC^2)}{n_{pos} \cdot n_{neg}}},$$

Want well defined confidence intervals. (DeLong's method? Anyone working on these things?)

radiology.143.1.7063747

Thinking about Weights

Paper in Construction

We have trained the LDA models on a weighted sample but without the weights (sklearn's **Linear** Discriminant Analysis (LDA) model, does not handle weights in training)
(We have plotted the ROC curves and model outputs with the weighs)

Can we do this though?
Is it stable?
Can we train on weighted (MC), test on unweighted (Data)?
Should we use the weights in training?
Do we want to capture the true p_T distribution in training (use the weights)
or prefer that the network learns uniformly across p_T bins (no weights)?

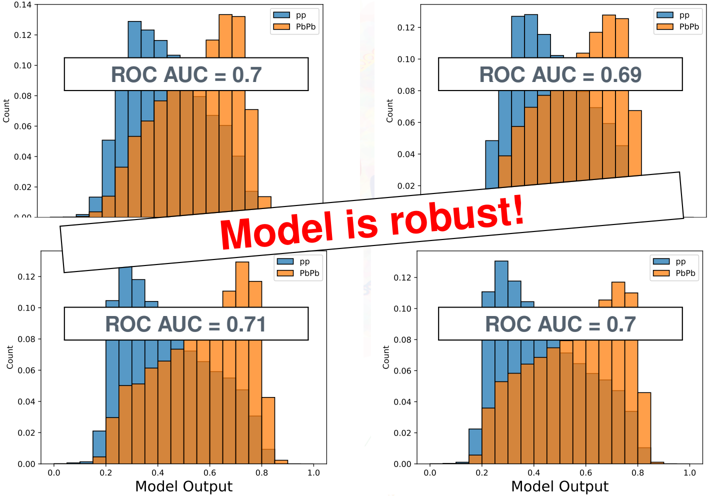
Is the model robust to this?

Thinking about Weights

Paper in Construction

Tested on: Weighted | Unweighted

Trained on: Weighted | Unweighted



Model Comparison (Supervised) No UE

Model	AUC	$1/\epsilon_{PbPb}$ at $\epsilon_{pp} = 50\%$
LDA EFP	0.8675	Y.YY
LDA EFP Ext.	0.9234	Y.YY
LDA NSub	0.8314	Y.YY
DNN EFP	0.9067	Y.YY
DNN EFP Ext.	**0.9420**	Y.YY
DNN NSub	0.9232	Y.YY

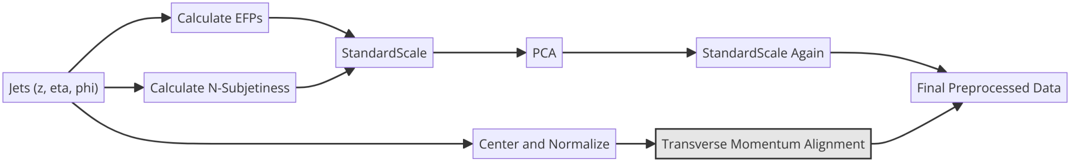
Model Comparison (Supervised) w/i UE Contamination

Model	AUC	$1/\epsilon_{PbPb}$ at $\epsilon_{pp} = 50\%$
LDA EFP	0.6964	Y.YY
LDA EFP Ext.	0.7132	Y.YY
LDA NSub	0.6900	Y.YY
DNN EFP	0.7142	Y.YY
DNN EFP Ext.	0.7176	Y.YY
DNN NSub	0.7075	Y.YY
EFN	0.7651 +/- 0.0004	Y.YY
EFN + EFP	0.8050 +/- 0.0002	Y.YY
EFN + EFP Ext.	0.8104 +/- 0.0004	Y.YY
EFN + NSub	0.8053 +/- 0.0003	Y.YY
EFN + EFP + NSub	0.8206 +/- 0.0001	Y.YY
EFN + EFP Ext. + NSub	**0.8265 +/- 0.0011**	Y.YY

Model Comparison MEFN w/i UE Contamination

Model	AUC	$1/\epsilon_{PbPb}$ at $\epsilon_{pp} = 50\%$
MEFN k=1, L=1024	0.7634 +/- 0.0007	Y.YY
MEFN k=2, L=32	0.7632 +/- 0.0006	Y.YY
MEFN k=1, L=512	0.7615 +/- 0.0008	Y.YY
MEFN k=3, L=16	0.7624 +/- 0.0004	Y.YY
MEFN k=2, L=64	0.7623 +/- 0.0002	Y.YY
MEFN k=4, L=16	0.7623 +/- 0.0001	Y.YY
MEFN k=2, L=16	0.7521 +/- 0.0018	Y.YY
MEFN k=1, L=256	0.7548 +/- 0.0013	Y.YY
MEFN k=1, L=128	0.7485 +/- 0.0017	Y.YY
MEFN k=1, L=64	0.7421 +/- 0.0017	Y.YY
MEFN k=1, L=32	0.7359 +/- 0.0022	Y.YY
MEFN k=1, L=16	0.7293 +/- 0.0086	Y.YY
MEFN k=2, L=8	0.7262 +/- 0.0050	Y.YY
MEFN k=1, L=8	0.7196 +/- 0.0031	Y.YY
MEFN k=3, L=3	0.7177 +/- 0.0065	Y.YY
MEFN k=4, L=4	0.7212 +/- 0.0033	Y.YY
MEFN k=4, L=3	0.7201 +/- 0.0025	Y.YY
MEFN k=8, L=3	0.7193 +/- 0.0020	Y.YY
MEFN k=3, L=2	0.7146 +/- 0.0020	Y.YY
MEFN k=4, L=2	0.7142 +/- 0.0020	Y.YY
MEFN k=16, L=2	0.7131 +/- 0.0020	Y.YY

Preprocessing Pipeline



Model Architectures

Model	Layers	Activation	Patience	Dropout	L2
DNN	(2048)	ReLU	30	0.2	0.005
EFN	Φ: (100, 100, 126), *F*: (100, 100, 100)	ReLU	30	0.075	-
EFN + Obs.	Φ: (100, 100, 126), *F*: (100, 100, 100)	ReLU	30	0.2	-
MEFN	Φ: (100, 100, L), *F*: (100, 100, 100)	ReLU	30	0.2	-

Observables

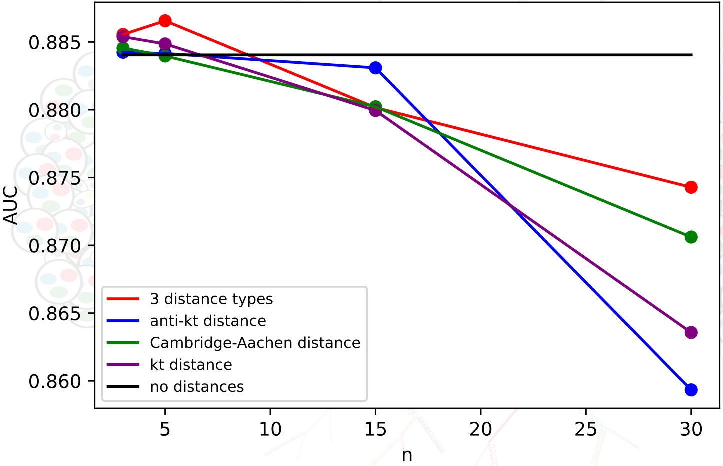
$$K = 21$$

$$\left\{\tau_1^{(1/2)},\ \tau_1^{(1)},\ \tau_1^{(2)}, \tau_2^{(1/2)},\ \tau_2^{(1)},\ \tau_2^{(2)},\ \ldots,\ \tau_{K-2}^{(1/2)},\ \tau_{K-2}^{(1)},\ \tau_{K-2}^{(2)}, \tau_{K-1}^{(1/2)},\ \tau_{K-1}^{(1)}\right\}$$

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

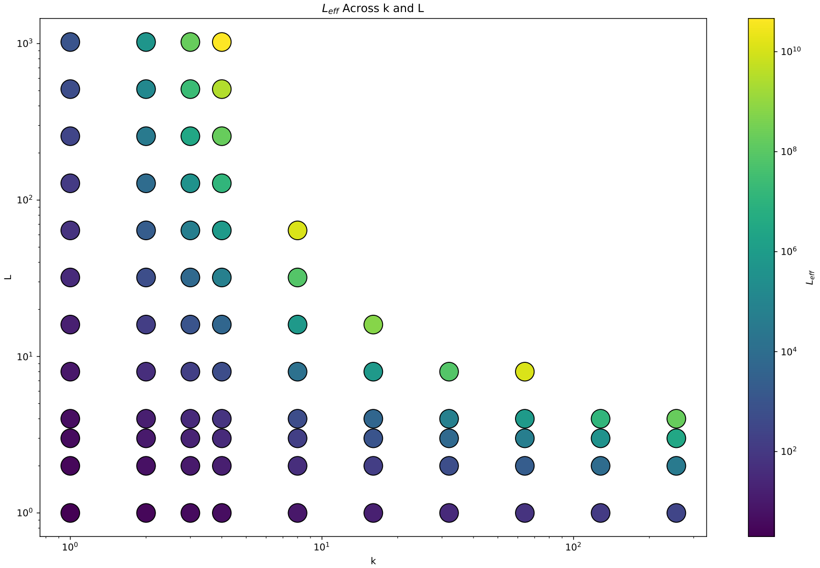
$$(\kappa, \beta) = ([0.5, 1], [0.5, 1, 2])$$

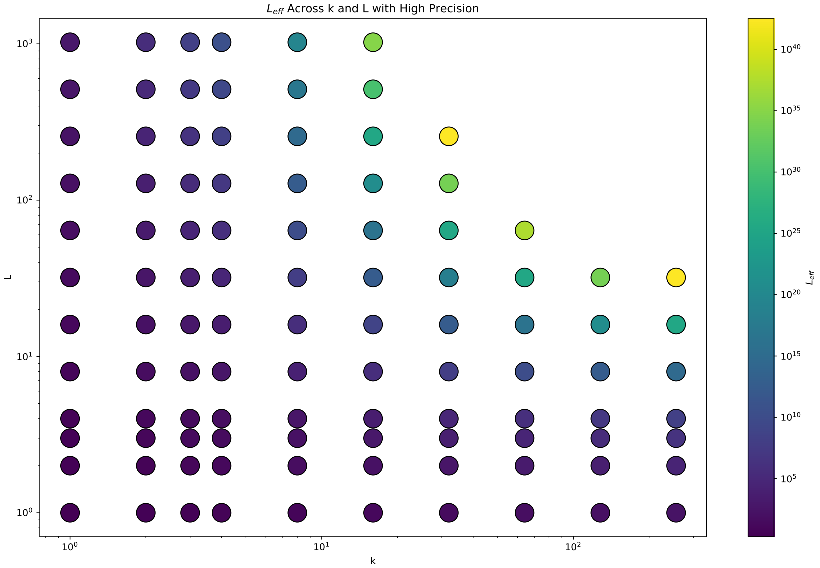
Adding distances to EFNs for quark vs gluon



By: Martim Pinto

Summer Student


L_{eff} Across k and L
L
k
L_{eff}
10^0
10^1
10^2
10^3
10^4
10^6
10^8
10^{10}


L_{eff} Across k and L with High Precision
L
k
10^0
10^1
10^2
10^3
10^{40}
10^{35}
10^{30}
10^{25}
10^{20}
10^{15}
10^{10}
10^5
L_{eff}