Learning to Identify Semi-Visible Jets

ML4JETS 2022

Related Paper: arxiv.org/abs/2208.10062

Taylor Faucett [1]
Daniel Whiteson [1]
Shih-Chieh Hsu [2]

[1] UC Irvine [2] U Washington

Taylor Faucett

Daniel Whiteson

Shih-Chieh Hsu

A Quick Overview of Interpretability and Jet Substructure

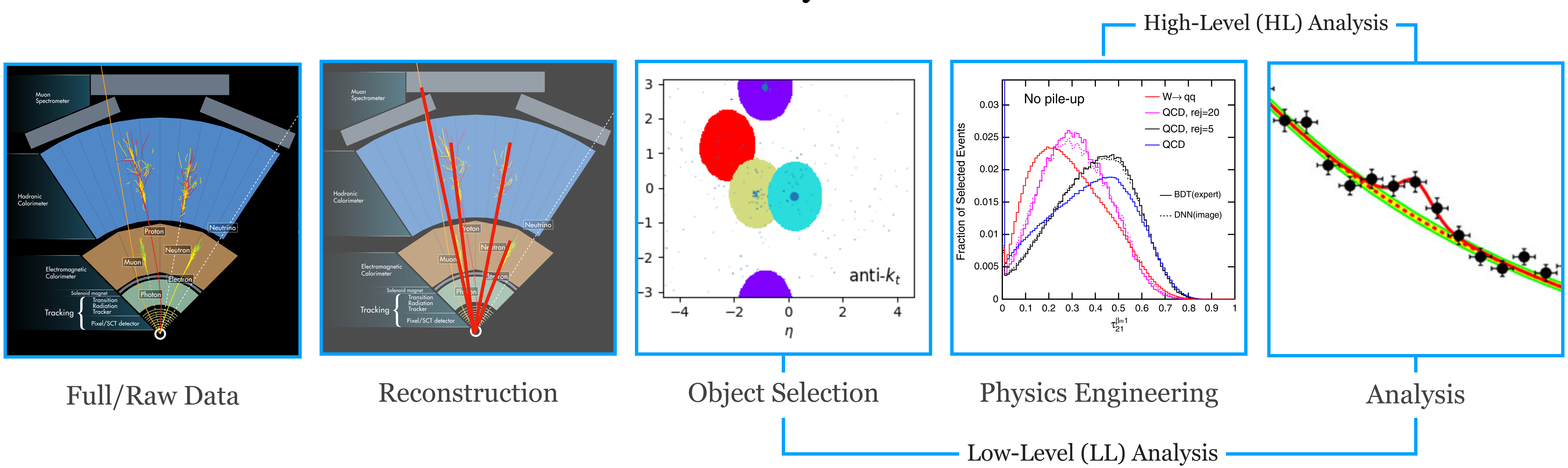
Data goes through a processing pipeline:

1. Raw Data
2. Reconstruction
3. Object Selection
4. "Physics Engineering"
5. Analysis

This has a few benefits

1. Features are physical/intelligible
2. Features are inspectable and can be validated
3. Performance Improves!

Dimensionality Reduction



- Jet: Collimated group of stable hadrons
- Form from “free” quarks/gluons which hadronize due to confinement
- Jets are detected as groups of particles in the calorimeter

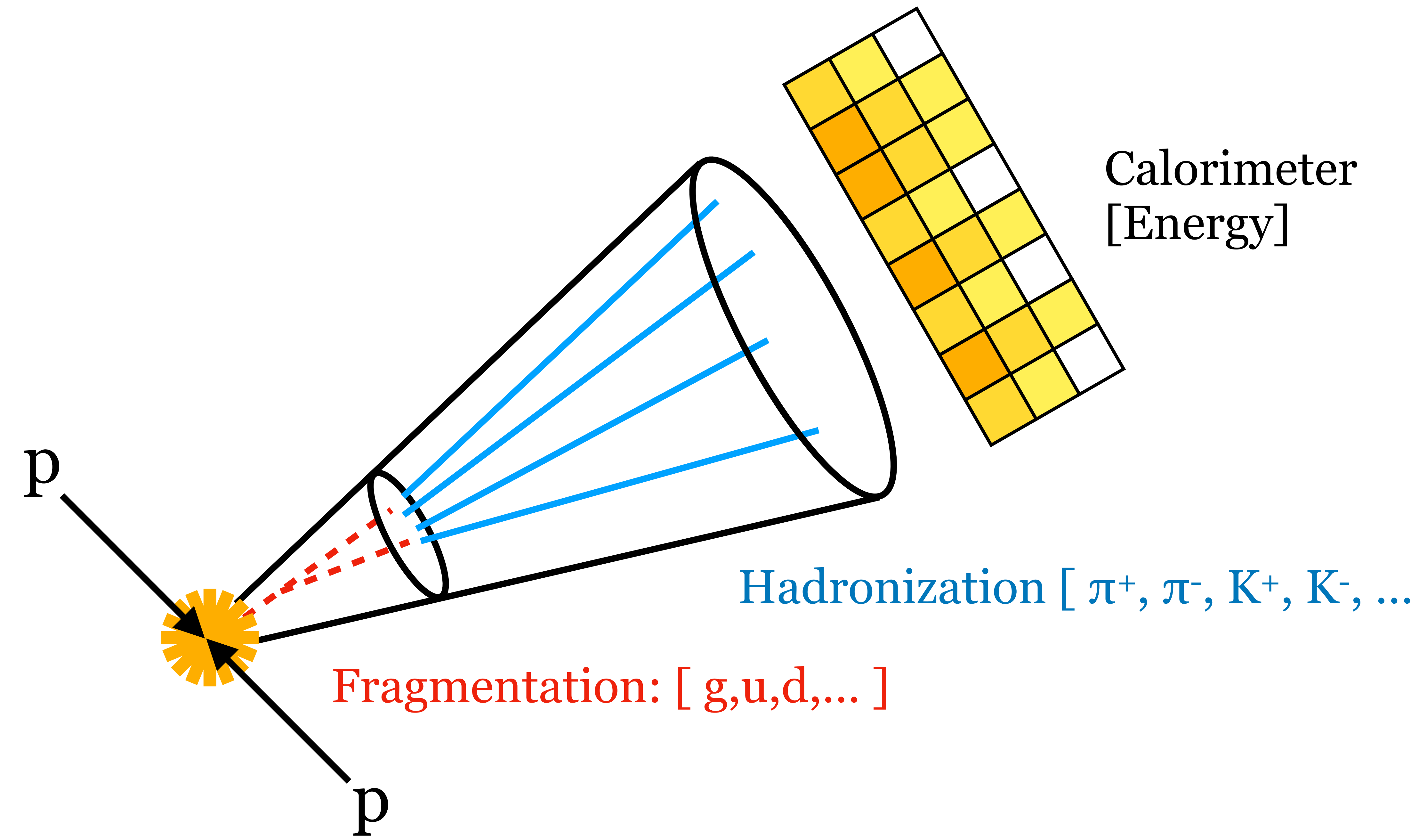
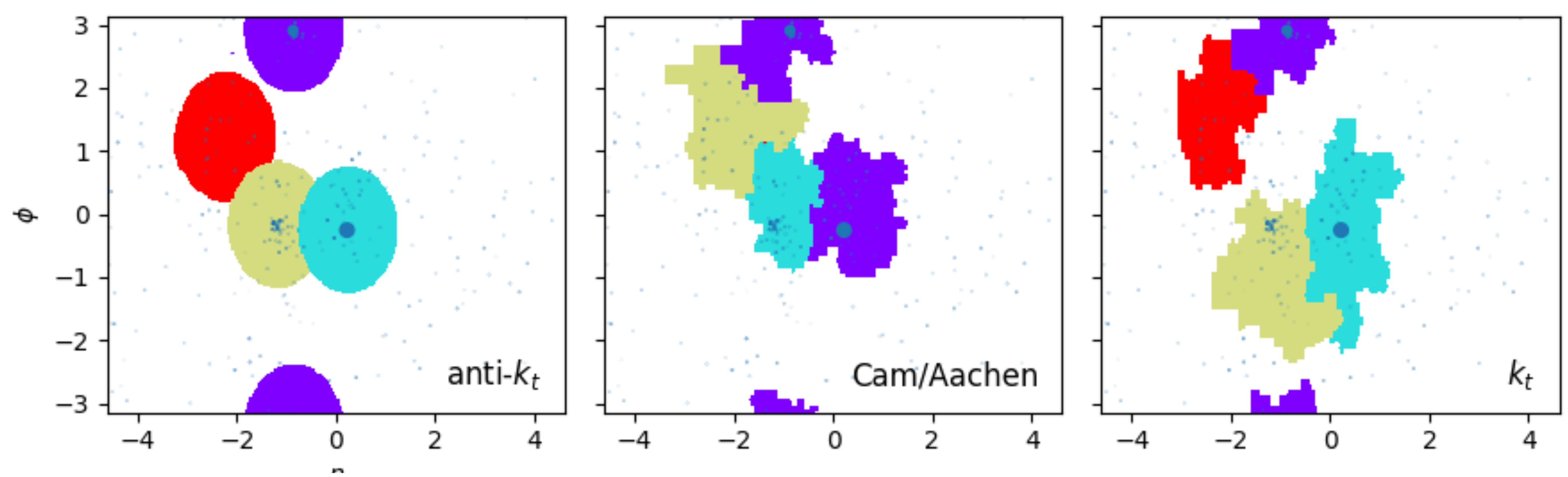


Diagram for quark/gluon hadronization



- Most Jet Substructure (JSS) observables are composed of jet constituents momentum fraction z_i and angular separation from the jet axis θ_i for a clustered jet with clustering radius R_0

$$z_i = \frac{p_{T,i}}{\sum p_T}, \qquad \theta_i = \frac{R_{i,\hat{n}}}{R_0}$$

- Quark/Gluon discrimination (Generalized Angularity)

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}}^{N} z_i^\kappa \theta_i^\beta$$

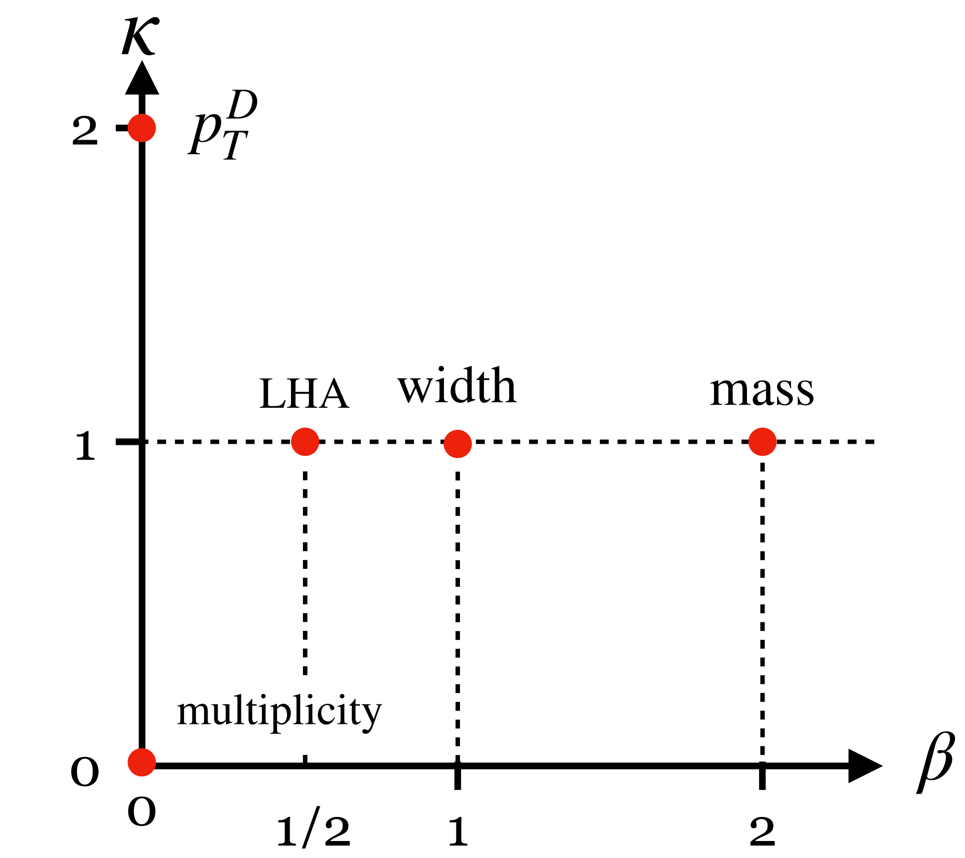
- W/Z/Higgs jets (Energy Correlation Functions = Higher Order GA)

$$ECF(2,\beta) = \sum_{i<j \in \text{jet}} p_{T,i} p_{T,j} \left(\theta_{ij}\right)^\beta$$

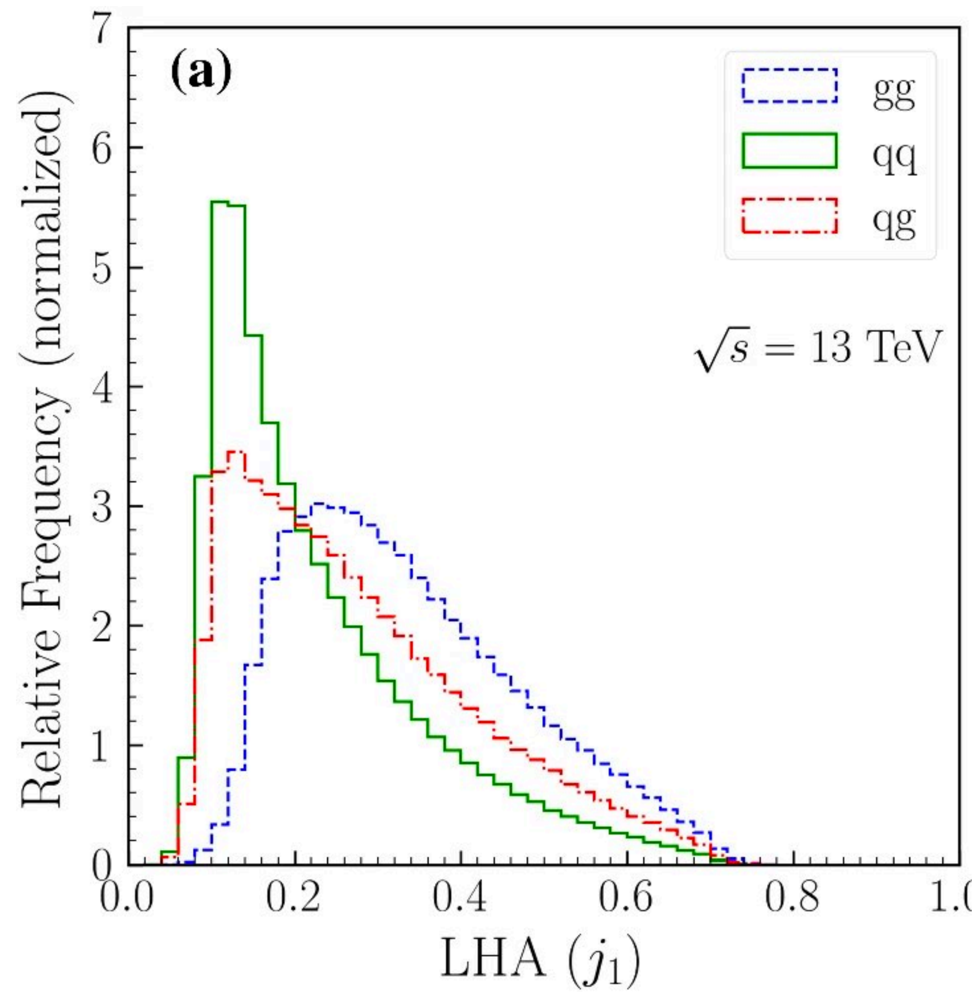
$$ECF(3,\beta) = \sum_{i<j<k \in \text{jet}} p_{T,i} p_{T,j} p_{T,k} \left(\theta_{ij}\theta_{jk}\theta_{ik}\right)^\beta$$

- N-Prong sensitivity (N-subjettiness)

$$\tau_N = \frac{1}{d_0} \sum_{i in \text{jet}}^{N} p_{T,i} \min\left\{\Delta\theta_{1,i}, \ldots, \Delta\theta_{N,i}\right\}, \text{where } d_0 = \sum_j p_{T,j} R_0$$

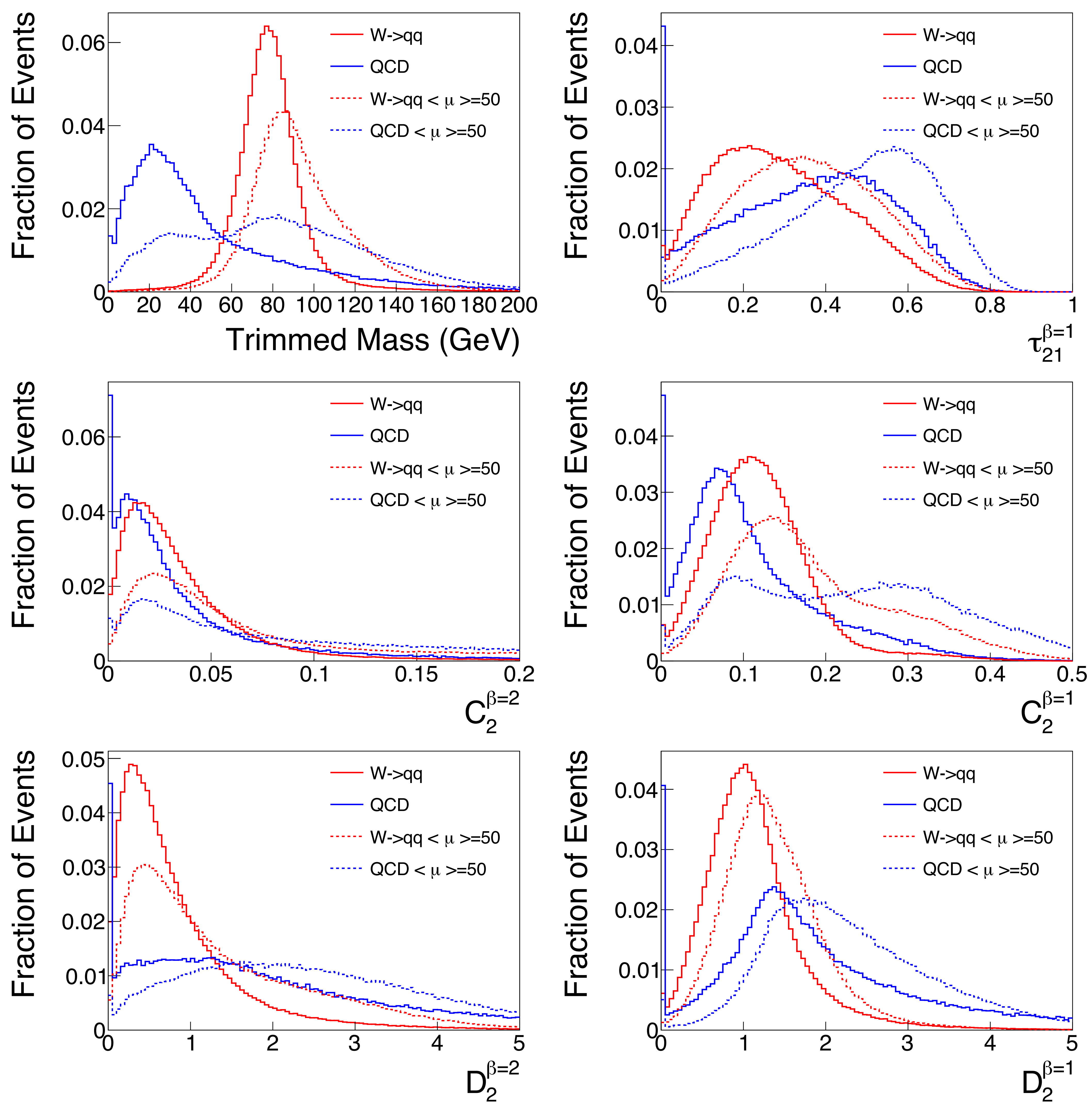


Generalized Angularity parameters



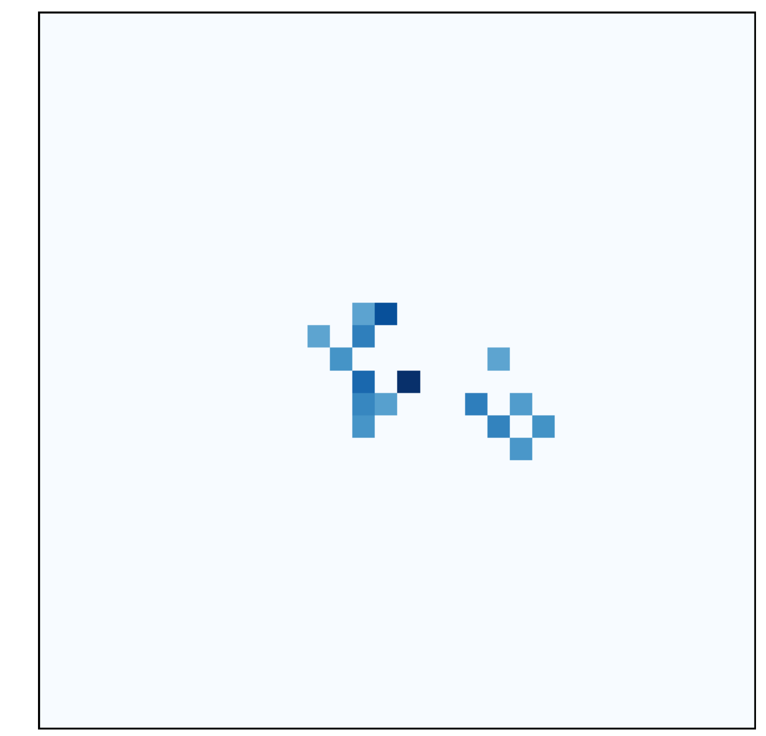
Les Houches Angularity (LHA) for quark/gluon discrimination

6 HL Variables

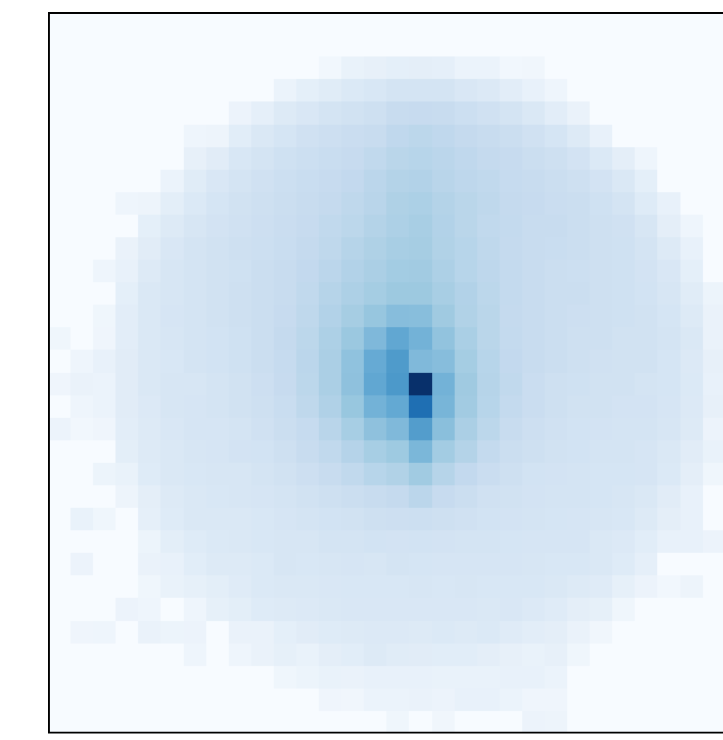


VS

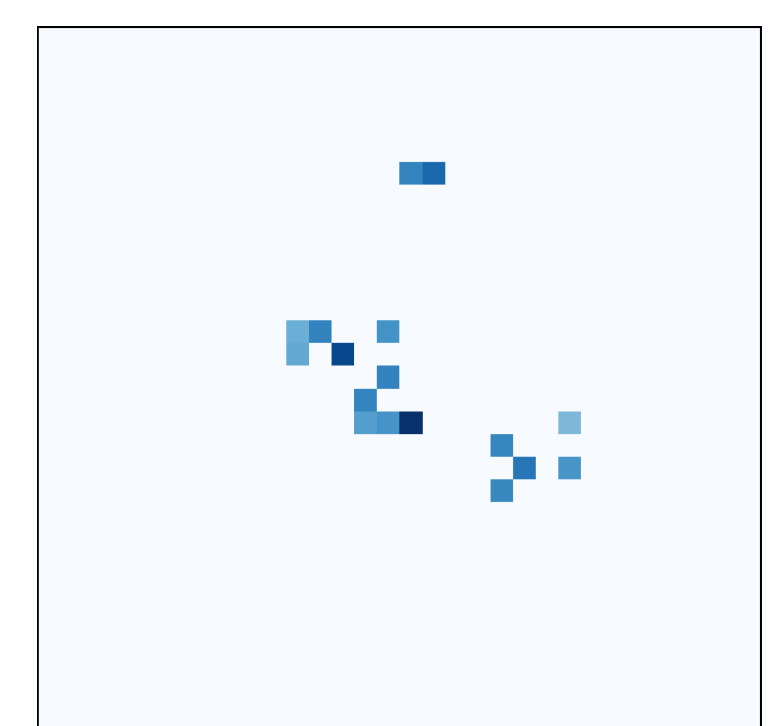
LL Jet Images



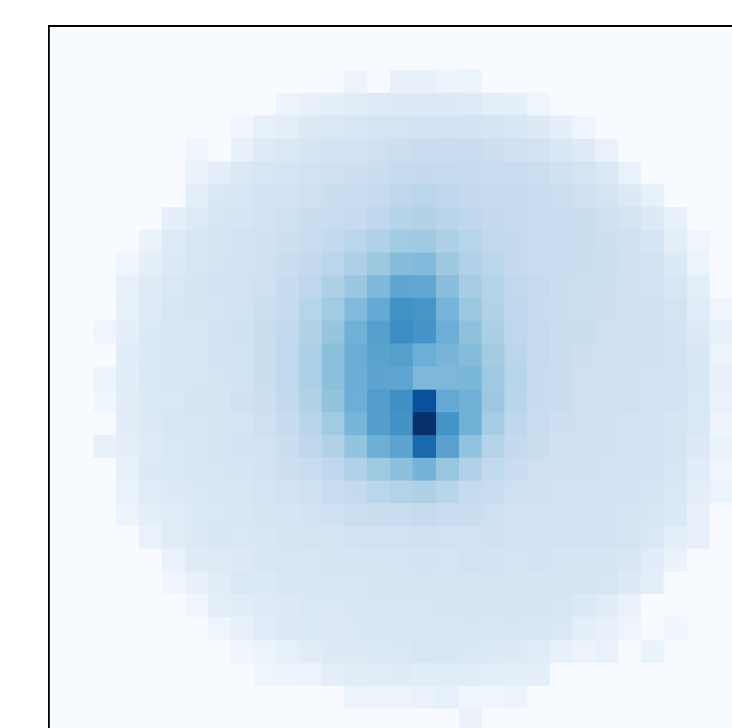
1 Event



Average of all events



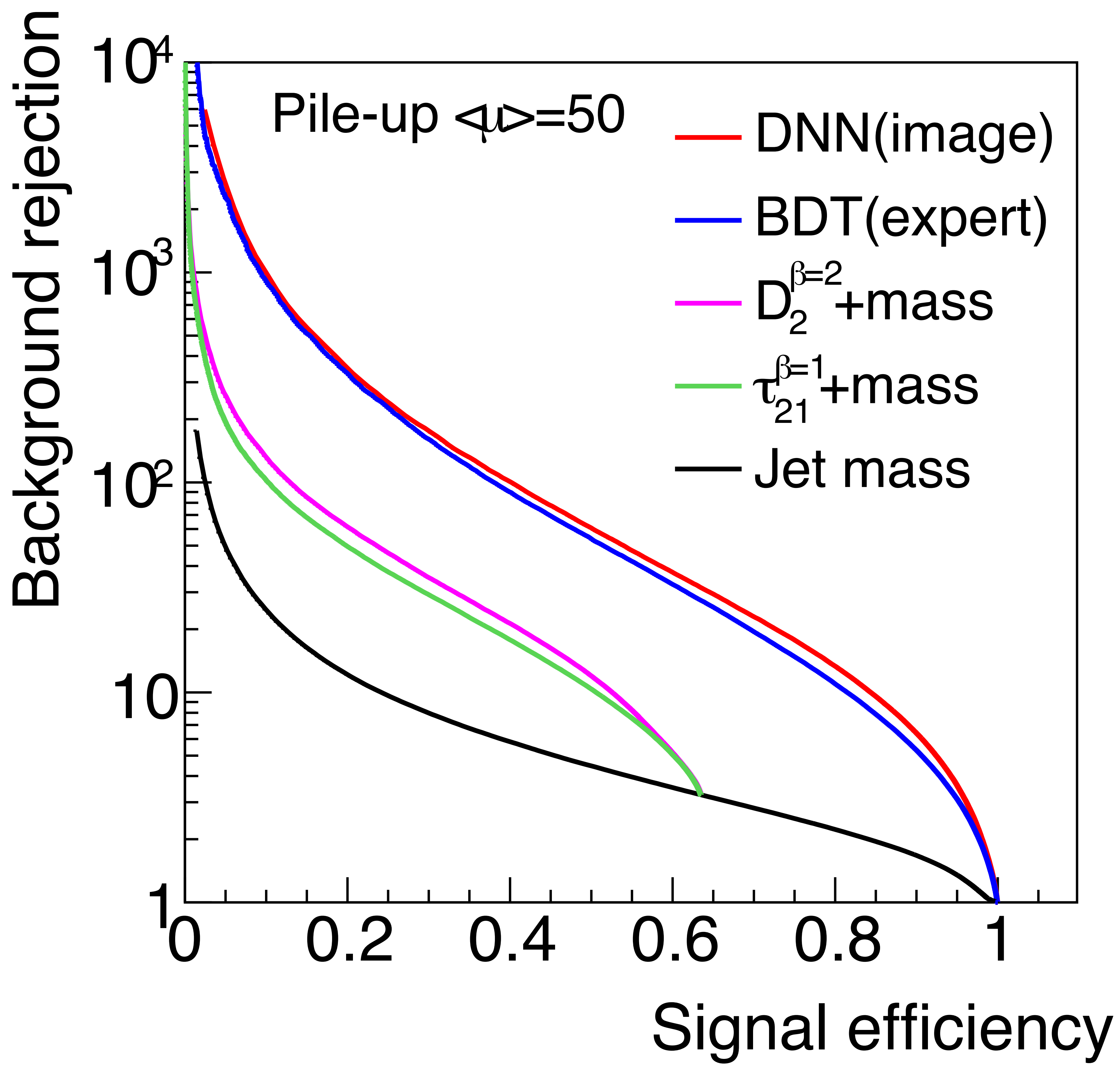
1 Event



Average of all events

- Baldi et al. find a CNN on jet images performs better than Jet Substructure
  - **Jet Images (red line)**: AUC = 95.30% ± 0.02%
  - **JSS (blue line)**: AUC = 95.00% ± 0.02%
- Where is that extra information coming from?
- Why don't our standard Jet Substructure observables contain this information?
- Is it real physics that we don't know about yet?

We've used a black box, so now what? ¯_(ツ)_/¯



Baldi, P., Bauer, K., Eng, C., Sadowski, P., & Whiteson, D. (2016, March 30). Jet Substructure Classification in High-Energy Physics with Deep Neural Networks. arXiv.org. http://doi.org/10.1103/PhysRevD.93.094034

This is a common situation which we have proposed a solution to

- SM Jets: “Mapping Machine-Learned Physics into a Human-Readable Space”
  - https://arxiv.org/pdf/2010.11998
- Muon Decay Jets: “Learning to Isolate Muons”
  - https://arxiv.org/pdf/2102.02278
- Electron vs Jet Delineation: “Learning to Identify Electrons”
  - https://arxiv.org/pdf/2011.01984

Will it work for Semi-Visible jets?

- Semi-Visible Jets: “Learning to Identify Semi-Visible Jets”
  - https://arxiv.org/pdf/2208.10062

Spoiler Alert! Yes it works but with some interesting caveats unique to semi-visible jets.

Solving The Problem

Energy Flow Polynomials (EFP): Complete linear basis set for jet substructure

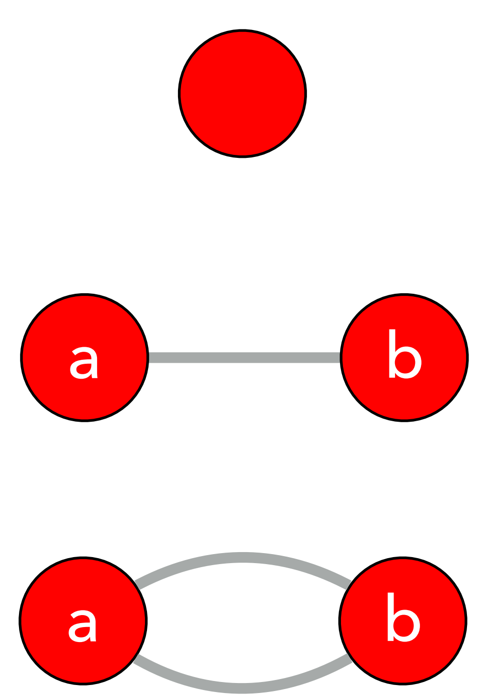
The set of EFPs is defined as all isomorphic graphs, with p_T and position (θ) as defined below

For every set of graphs, we can also modify 2 parameters (κ,β)

$$z_i = \frac{p_{T,i}^{\kappa}}{\sum_i p_{T,i}}$$

$$\theta_{ij} = \left(\Delta y_{ij}^2 + \Delta_{ij}^2\right)^{\beta/2}$$

Graph components



Node/Vertex: $= \sum_a^N z_a$

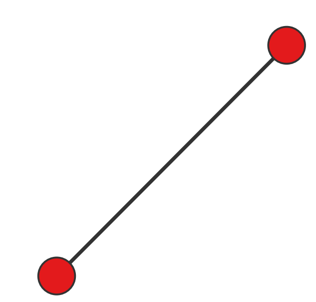
Edges: $= \theta_{ab}$

Multiple Edges $= \left(\theta_{ab}\right)^2$

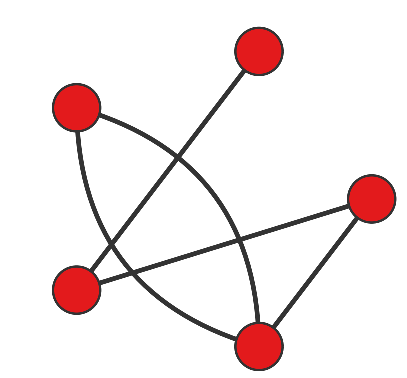
$$z_a = p_{T,a}^{\kappa}$$

$$\theta_{ab} = \left(\Delta\eta_{ab}^2 + \Delta\varphi_{ab}^2\right)^{\beta/2}$$

Examples



$$= \sum_a^N \sum_b^N z_a z_b \theta_{ab}$$



$$= \sum_a^N \sum_b^N \sum_c^N \sum_d^N \sum_e^N z_a z_b z_c z_d z_e \theta_{ac}^2 \theta_{bd} \theta_{be} \theta_{cd}$$

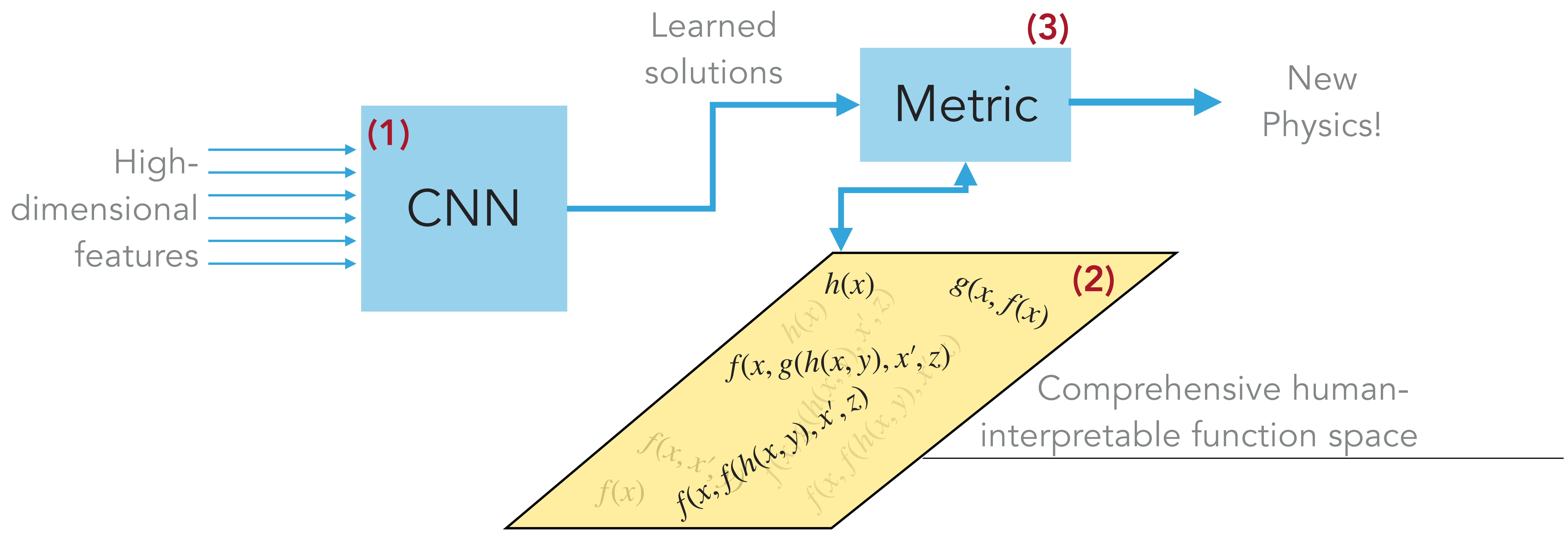
Komiske, P. T., Metodiev, E. M., & Thaler, J. (2017, December 19). Energy flow polynomials: A complete linear basis for jet substructure.

- Can we make a model that is made up entirely of intelligible high-level features but which is "equivalent" in its decision making to the CNN? A LL solution that performs better than the HL
- The parts we need are:
    - We have this from our CNN on jet images.

  (1) A LL but powerful solution

  (2) A "human readable" space of HL variables.

  (3) A metric for mapping the LL solution into the HL features.



We want an equivalent to ROC for 2 discriminating functions [f(x) and g(x)]. Classification decision of two functions at different thresholds.

Step 1

For points from signal and background (x and x'), we compare how each function maps those points relative to one another.

$$\mathrm{DO}(x, x') = \Theta\left[(f(x) - f(x')) \cdot (g(x) - g(x'))\right]$$

Step 2

Sum over all combinations of signal/ background decision orderings

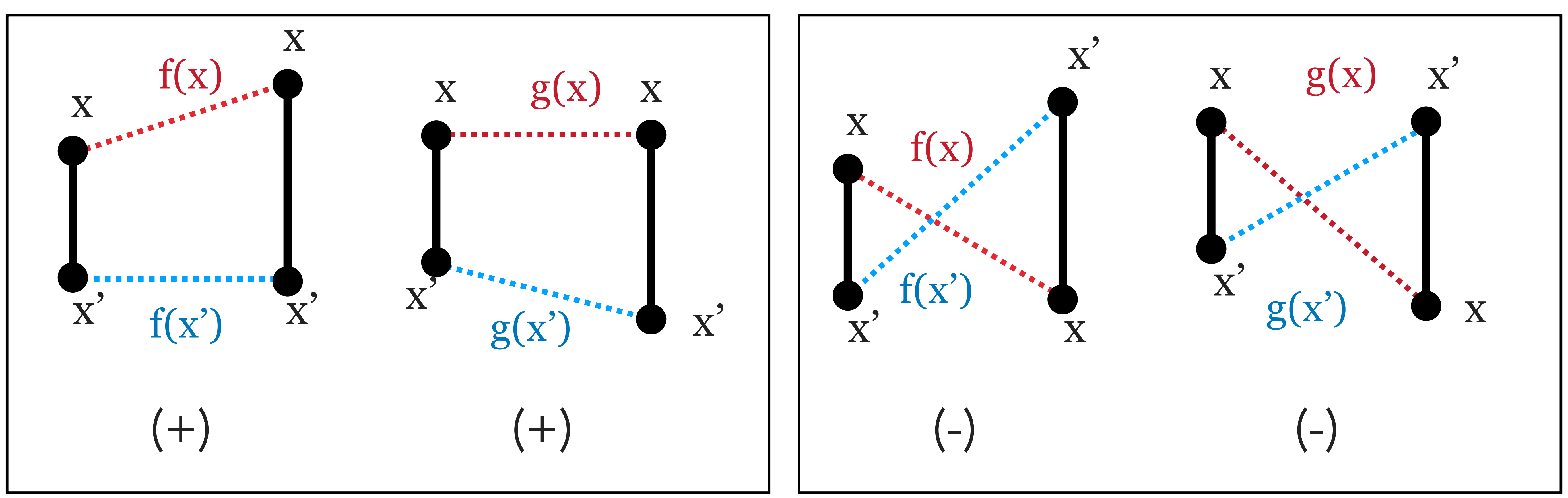
$$\mathrm{ADO}' = \sum DO(x, x')$$

Step 3

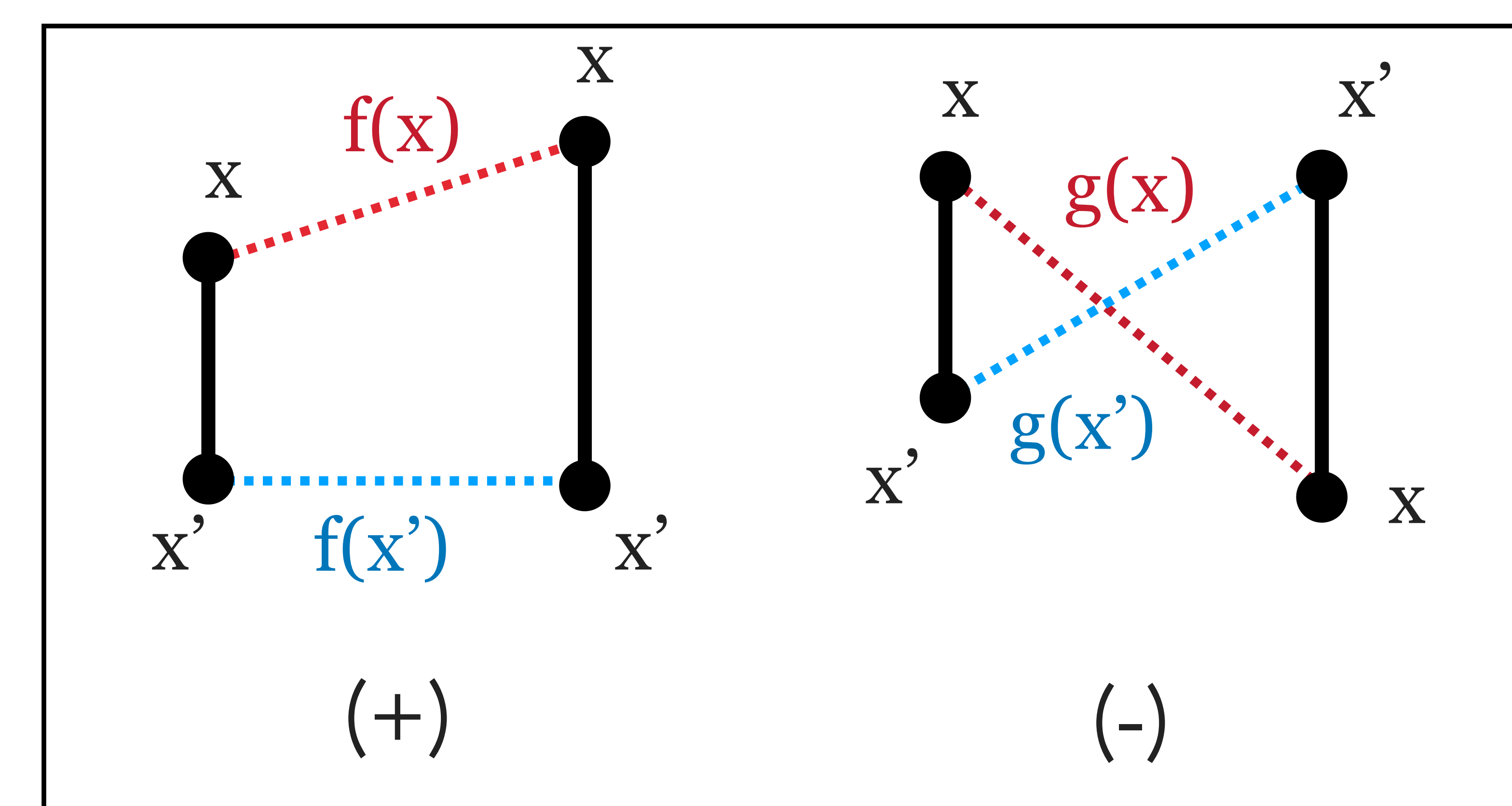
Consistent "dissimilarity" can be inverted to predict "similarity". Map all ADO less than 0.5

$$\mathrm{ADO} = 1 - \mathrm{ADO}_{<\frac{1}{2}}$$

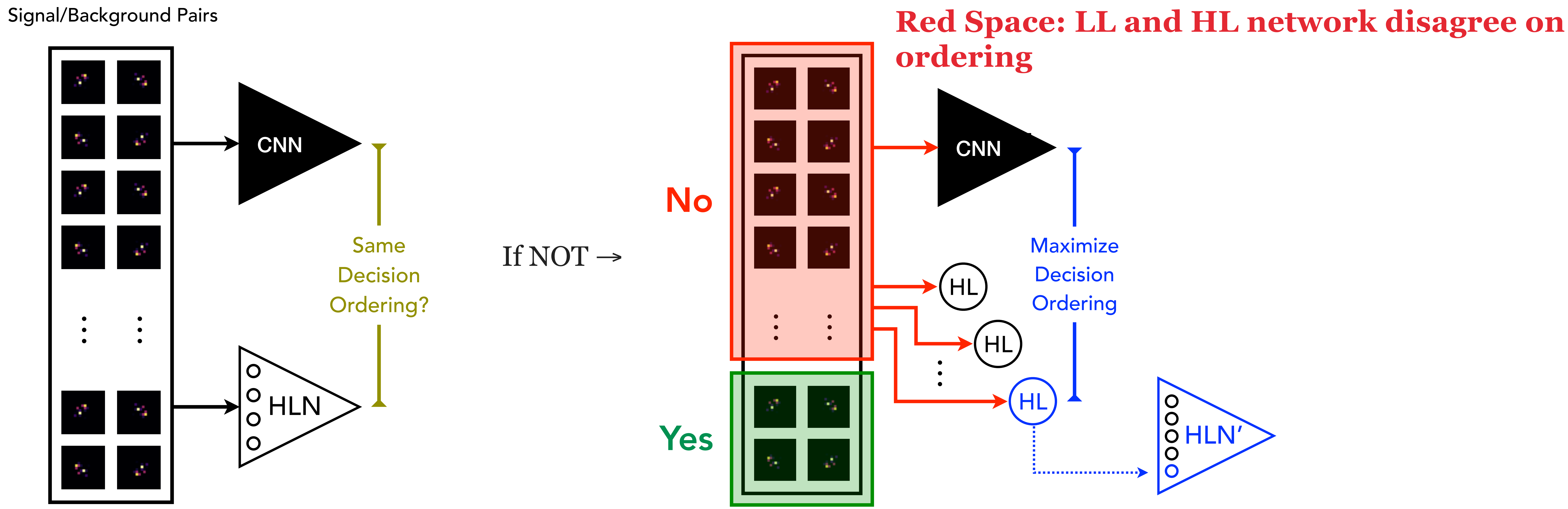
Similar Orderings



Dissimilar Ordering



We can compare NN decision making. Where does the HL network and LL network disagree?

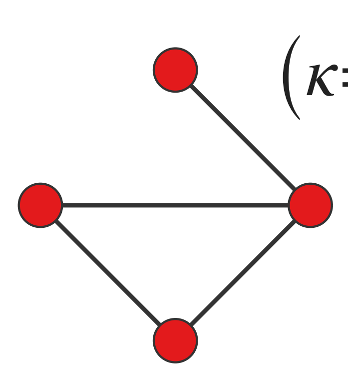


Use ADO to choose EFP that makes similar choices to the LL network in the "differently ordered" red space

We only need 1 new observable to achieve equal performance with the CNN!

Observable	AUC	ADO[CNN, Obs.]	
M_{jet}	0.898 ± 0.004	0.807	
$C_2^{\beta=1}$	0.660 ± 0.006	0.584	
$C_2^{\beta=2}$	0.604 ± 0.007	0.548	
$D_2^{\beta=1}$	0.790 ± 0.005	0.743	
$D_2^{\beta=2}$	0.807 ± 0.005	0.762	
$\tau_2^{\beta=1}$	0.662 ± 0.006	0.600	
6HL	0.9504 ± 0.0002	0.971	**Original HL**
CNN	0.9531 ± 0.0002	1.000	**LL network**
$7\text{HL}_{\text{black-box}}$	0.9528 ± 0.0003	0.971	**Original HL + 1 EFP**

Selected EFP from Guided Iteration

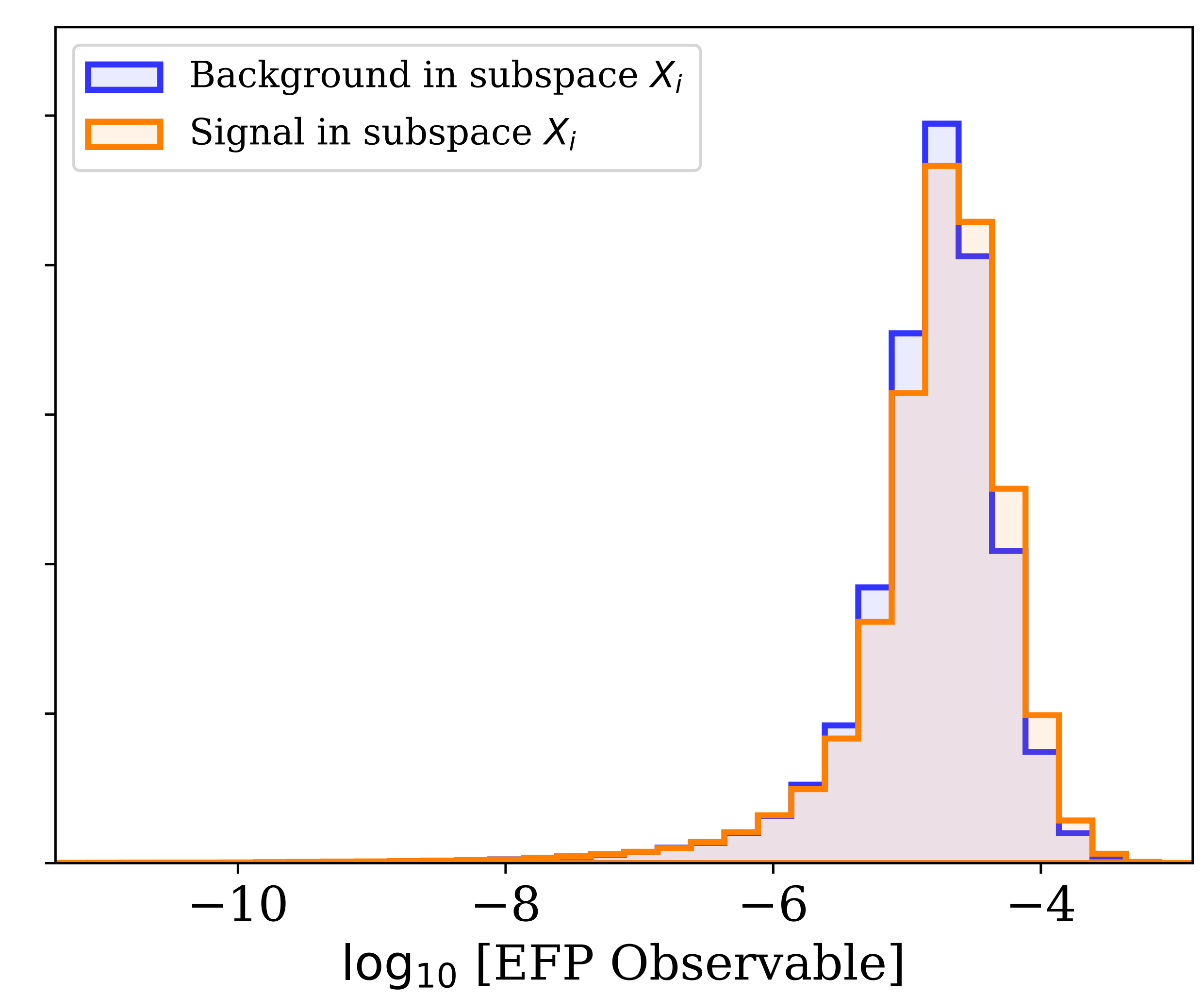


$(\kappa{=}2,\ \beta{=}1/2)$

$$= \sum_{a,b,c,d=1}^{N} z_a^2 z_b^2 z_c^2 z_d^2 \sqrt{\theta_{ab}\theta_{bc}\theta_{ac}\theta_{ad}}$$

Noteworthy details about the selected EFP

- Not Infrared-safe (k ≠ 1)
- β=1/2 is probing small-angle behaviour
- Chromatic #3 graph (probing deviations from 2-prong substructure)
- Chromatic Number = Minimum number of prongs to not vanish

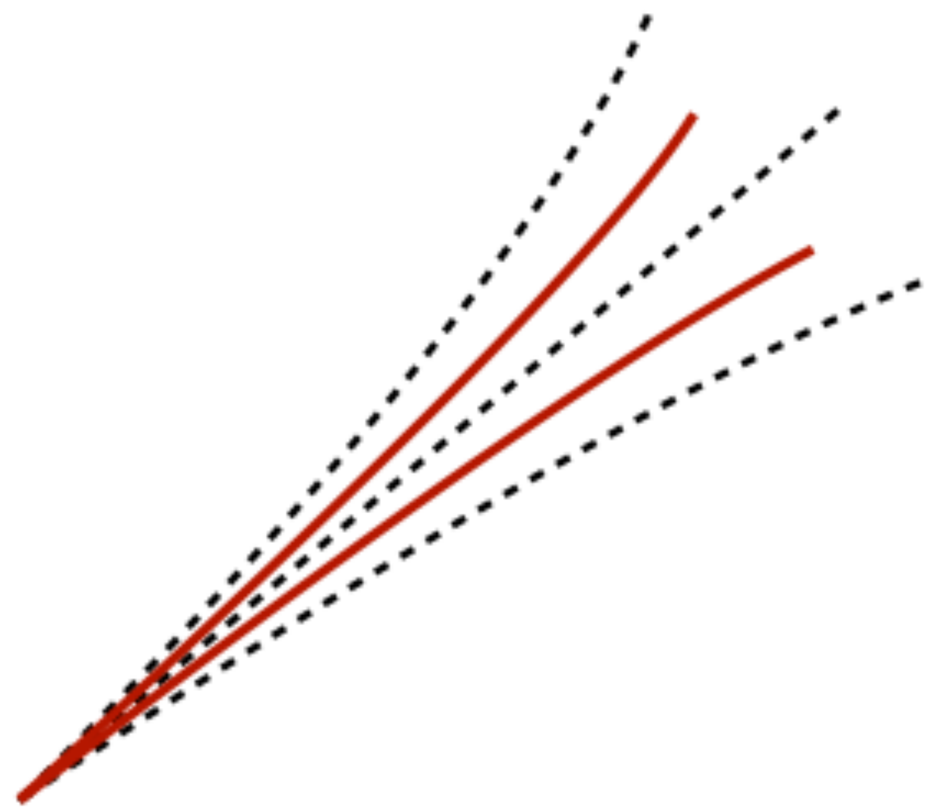


EFP Distribution (differently ordered points)

- What are Semi-Visible Jets?
  - SM process with partial decay mode into DM quarks/gluons.
  - Two popular examples described by Cohen, Lisanti, Lou (https://arxiv.org/pdf/1503.00009.pdf) involve an s-channel and t-channel process.
  - Decay into visible/invisible fraction is characterized by r_{inv} value
  - $r_{\text{inv}} = \left\langle \frac{\text{\# of stable dark hadrons}}{\text{\# of hadrons}} \right\rangle$

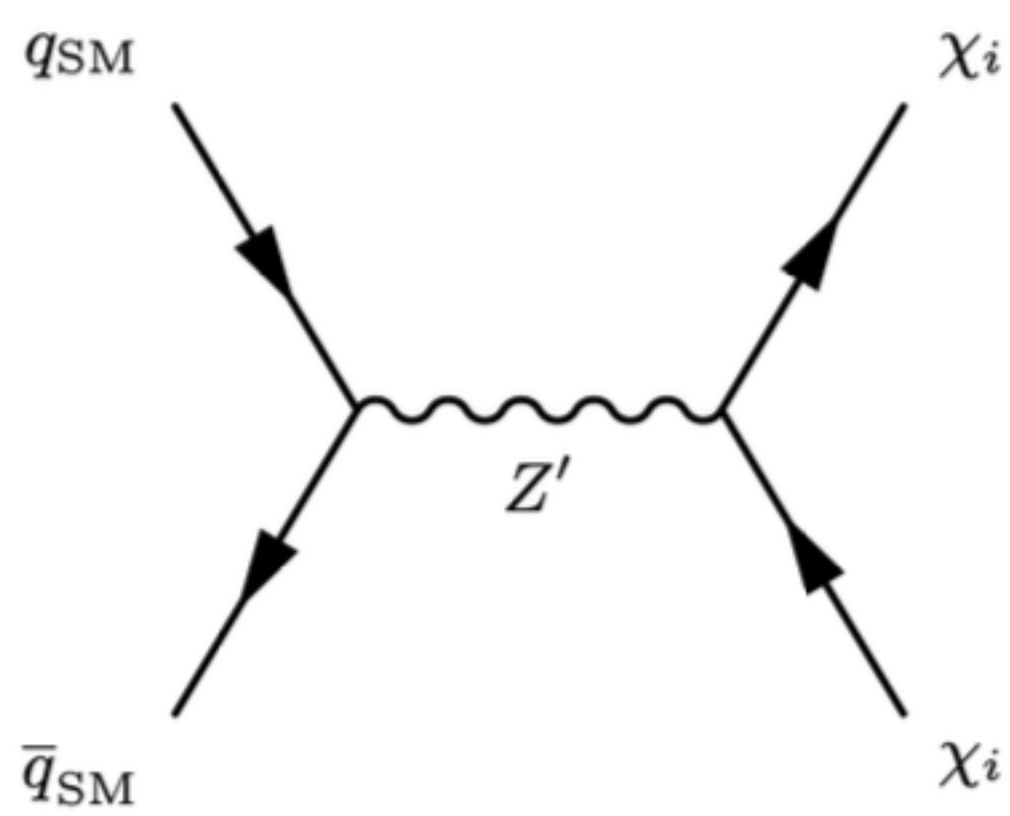
(a) Rapid Decay (Visible), $r_{\text{inv}} = 0$

(b) Stable Dark Jet (Invisible), $r_{\text{inv}} = 1$

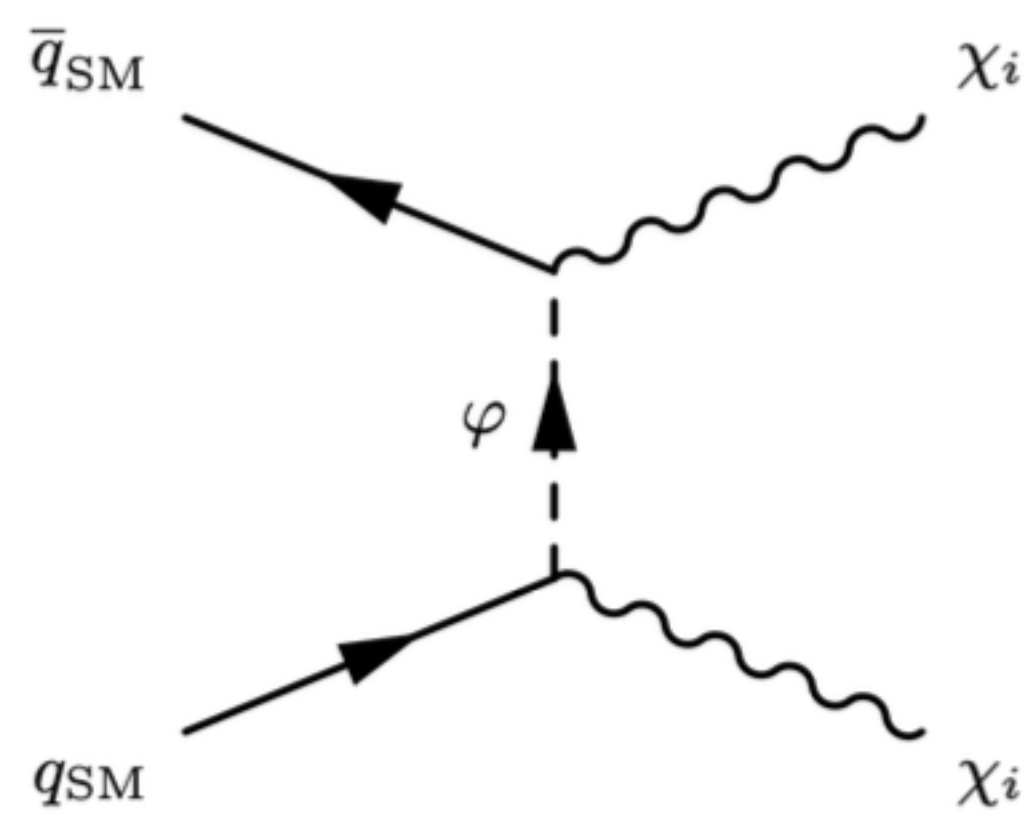


(c) Fractional Decay (Semi-Visible), $r_{\text{inv}} \in (0, 1)$

- Signal:
  - s-channel and t-channel SVJ
  - $pp \to Z' \to \chi_1\bar{\chi}_1$
  - $pp \to \varphi \to \chi_1\bar{\chi}_1$
  - COM Energy $\sqrt{s} = 13$,TeV
  - Hidden Valley settings (e.g. Mediator mass, Dark Quark mass, etc matches Cohen, Lisanti, Lou)
  - r_{inv} = [0.0, 0.3, 0.6]
- Background:
  - SM at the same COM energy: $pp \to jj$

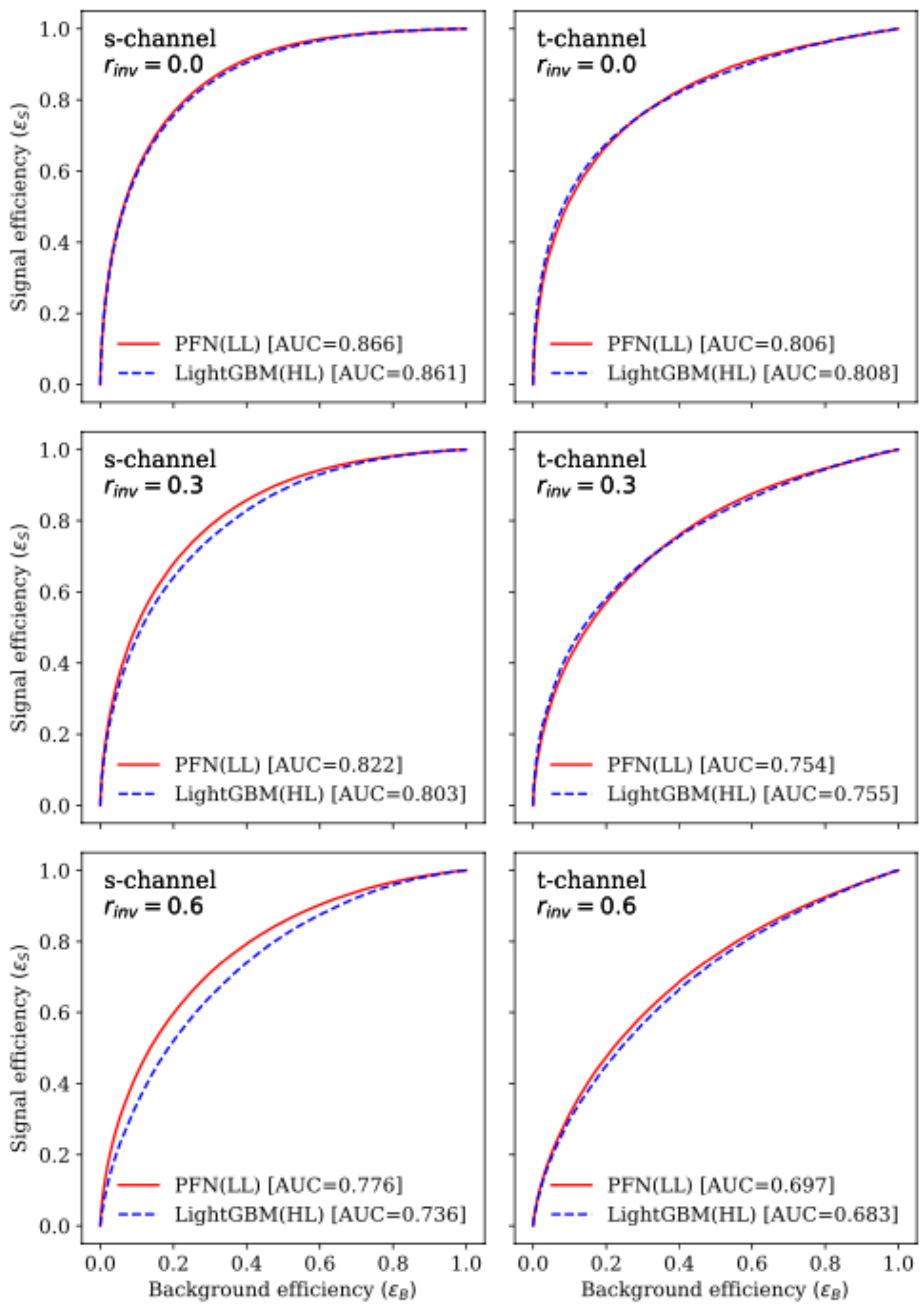


(a) s-channel



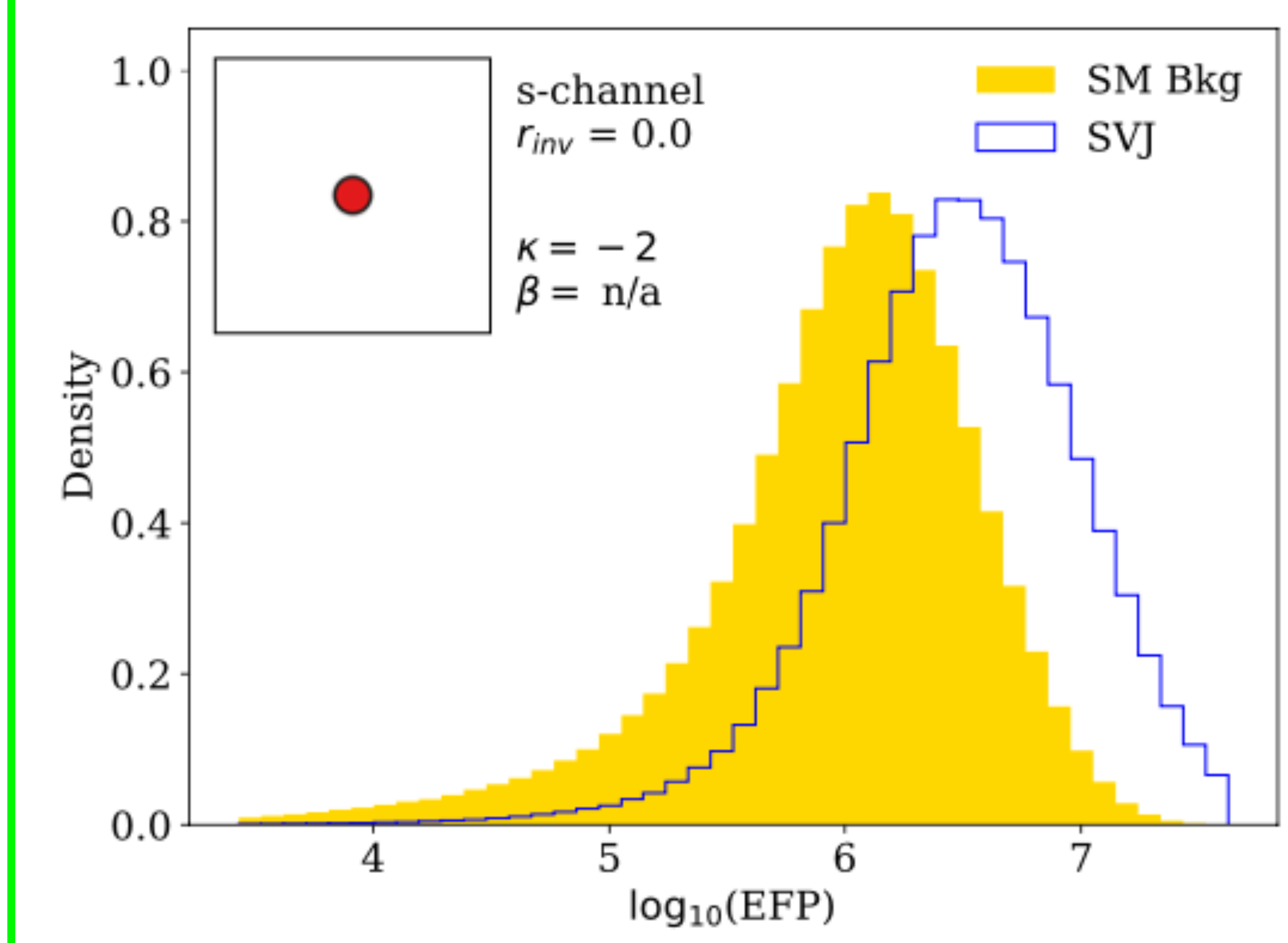
(b) t-channel

- LL trained with a Particle Flow Network using calorimeter constituents (p_T, η, ϕ)
- HL trained with a Boosted Decision Tree (LightGBM) on most common JSS features:
    - Jet pT
    - **Generalized Angularities:** LHA, pTD, Jet width, e_{mass}, multiplicity
    - **N-subjettiness:** $\tau_{21}^{\beta=1}$, $\tau_{32}^{\beta=1}$
    - **Energy Correlation Functions:** $C_2^{\beta=1}$, $C_2^{\beta=2}$, $D_2^{\beta=1}$, $D_2^{\beta=2}$, e_2, e_3
    - **Splitting Function:** z_g



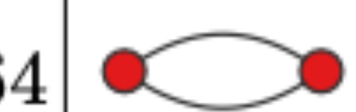
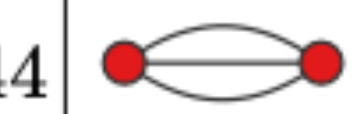
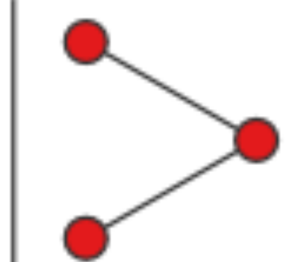
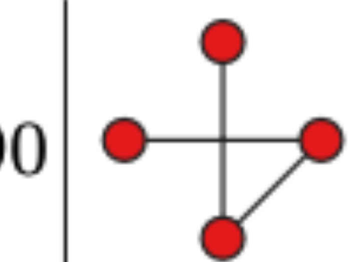
- EFP Space
  - All Graphs where: $d \leq 5$, $\kappa \in [2, -1, 0, 1/2, 1, 2, 4]$ and $\beta \in [1/10, 1/2, 1, 2, 4]$

Process	r_{inv}	HL network AUC, ADO[.,PFN]	HL+EFP network EFP	κ	β	HL+EFP network AUC, ADO[.,PFN]	PFN AUC
s-channel	0.0	0.861, 0.858		-2		0.864, 0.863	0.866
s-channel	0.3	0.803, 0.839		1	$\frac{1}{2}$	0.807, 0.840	0.822
s-channel	0.6	0.736, 0.818		-1	2	0.747, 0.821	0.776
t-channel	0.6	0.683, 0.787		-2	$\frac{1}{10}$	0.690, 0.792	0.697

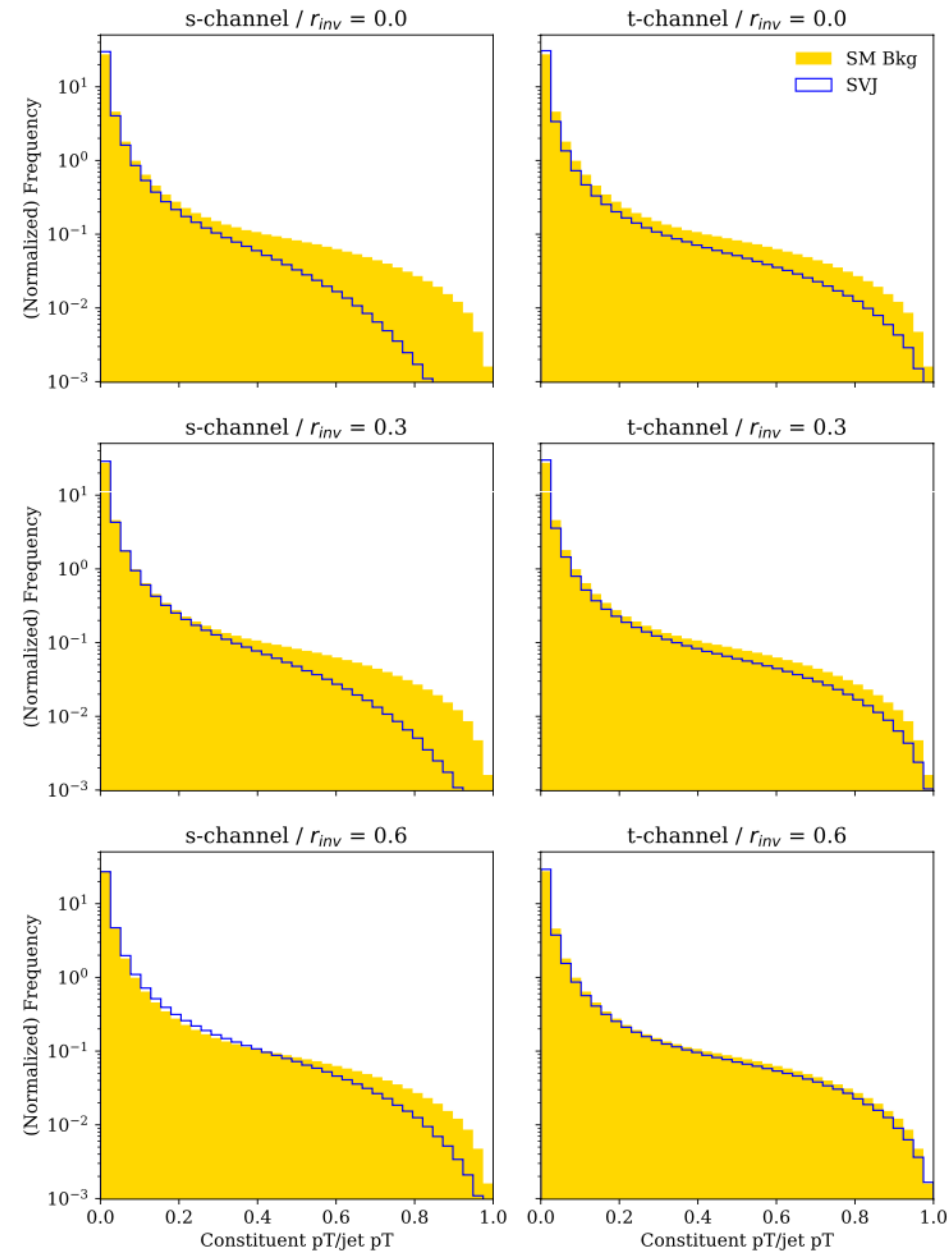


Is the Guided Search Failing Us?

- Two potential explanations:
  - The information exists as an EFP but the guided search is failing to find it.
  - The information doesn't exist in our EFPs
- To answer this question we perform a "greedy search":
  - Train a model for every combination of HL + EFP. Do any of these combinations fill the gap?

Process	r_{inv}	HL AUC	Pass 1 Graph	Pass 1 κ	Pass 1 β	Pass 1 AUC	Pass 2 Graph	Pass 2 κ	Pass 2 β	Pass 2 AUC	Guided HL AUC	PFN AUC
s-channel	0.0	0.861		$\frac{1}{2}$	2	0.864		2	$\frac{1}{10}$	0.866	0.864	0.866
s-channel	0.3	0.803		4	2	0.807		-1	1	0.809	0.807	0.822
s-channel	0.6	0.736		4	4	0.744		-2	$\frac{1}{10}$	0.747	0.747	0.776
t-channel	0.6	0.683		-1	$\frac{1}{10}$	0.690		-2	4	0.692	0.690	0.697

- We are doing "Guided Iteration" because it should help tell us what the model is learning.
- What do the selected EFPs tell us about the model?
- $\kappa < 0$ values are popular choices. This is sensitive to low-pT information.
- Data processing (i.e. jet clustering and pT cuts) uses a cut of constituents < 5% of the leading pT.
- What if we relax that cut and/or raise it?
- How does that impact model performance?



	s-channel								
	$r_{\text{inv}} = 0.0$			$r_{\text{inv}} = 0.3$			$r_{\text{inv}} = 0.6$		
f_{cut}	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap
0.00	0.908	0.895	0.013	0.853	0.829	0.024	0.788	0.739	0.049
0.05	0.866	0.861	0.005	0.822	0.803	0.019	0.776	0.736	0.040
0.10	0.847	0.848	-0.001	0.790	0.790	0.000	0.746	0.721	0.025
0.15	0.838	0.843	-0.005	0.784	0.785	-0.001	0.738	0.717	0.021

	t-channel								
	$r_{\text{inv}} = 0.0$			$r_{\text{inv}} = 0.3$			$r_{\text{inv}} = 0.6$		
f_{cut}	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap
0.00	0.825	0.817	0.008	0.748	0.737	0.011	0.662	0.647	0.0015
0.05	0.806	0.808	-0.002	0.754	0.755	-0.001	0.697	0.683	0.014
0.10	0.741	0.742	-0.001	0.662	0.663	-0.001	0.595	0.597	-0.002
0.15	0.731	0.740	-0.009	0.655	0.661	-0.006	0.593	0.596	-0.003

- This model of Semi-Visible Jets are not well described by standard JSS observables.
- Unlike many other jet classification tasks, they are also not particularly well-described by a compact set of new observables in the EFP space.
- SVJ classification appears particularly sensitive to low-pt/soft-emission constituents and generally IRC unsafe observables.
- However, even after aggressive trimming of low-pT information some LL-HL gaps remain.
- Although Low-Level networks can capture this information, they do so in a way we can't currently explain or validate.
- These results strongly motivate a need to understand the low-pT dependence for SVJ events AND a better space than EFPs to search for observables that capture the remaining constituents.

Questions ?