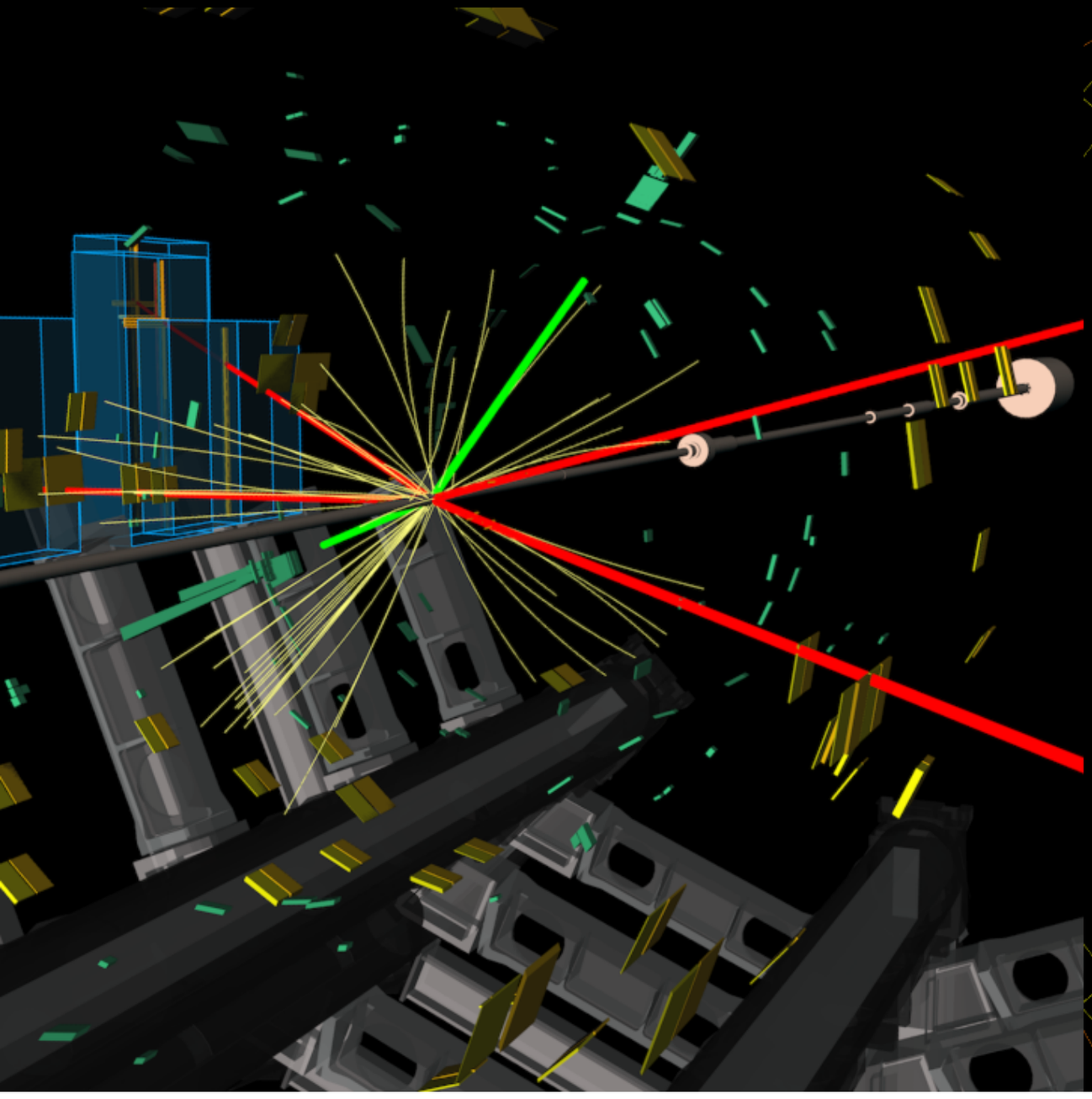
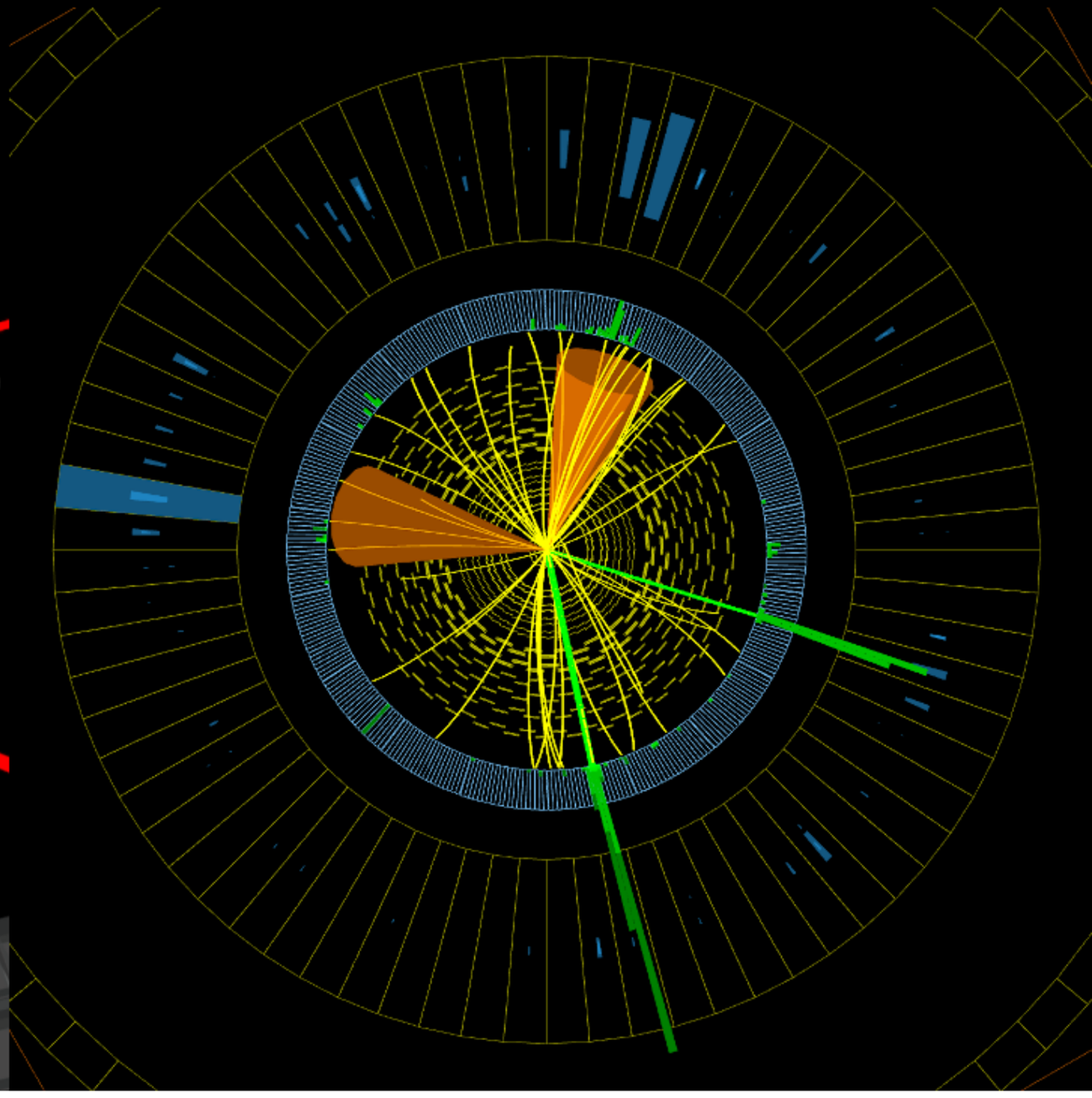
Jet clustering techniques at the LHC





Overview of this talk

1. Data and its acquisition

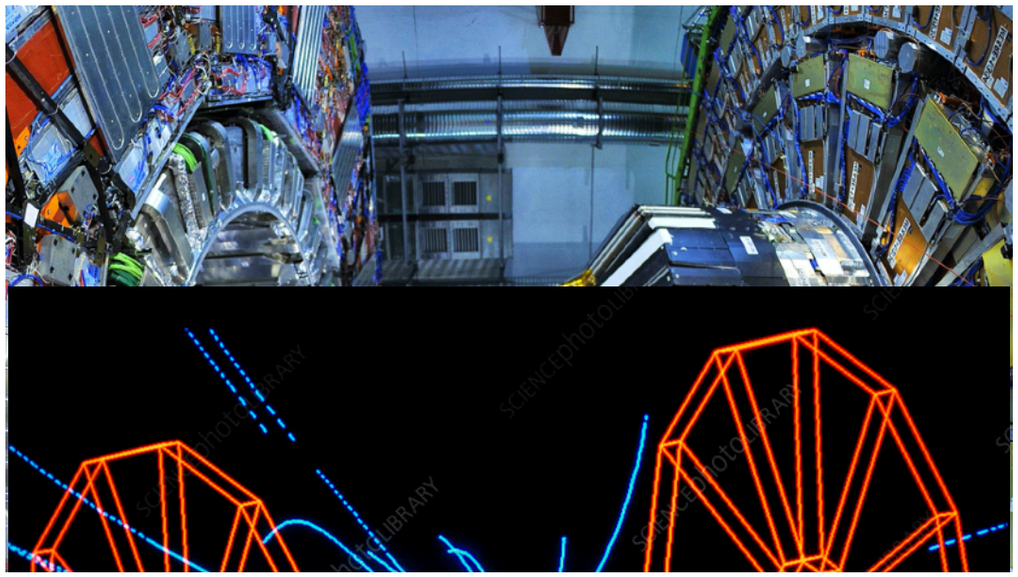
2. The ML problem statement

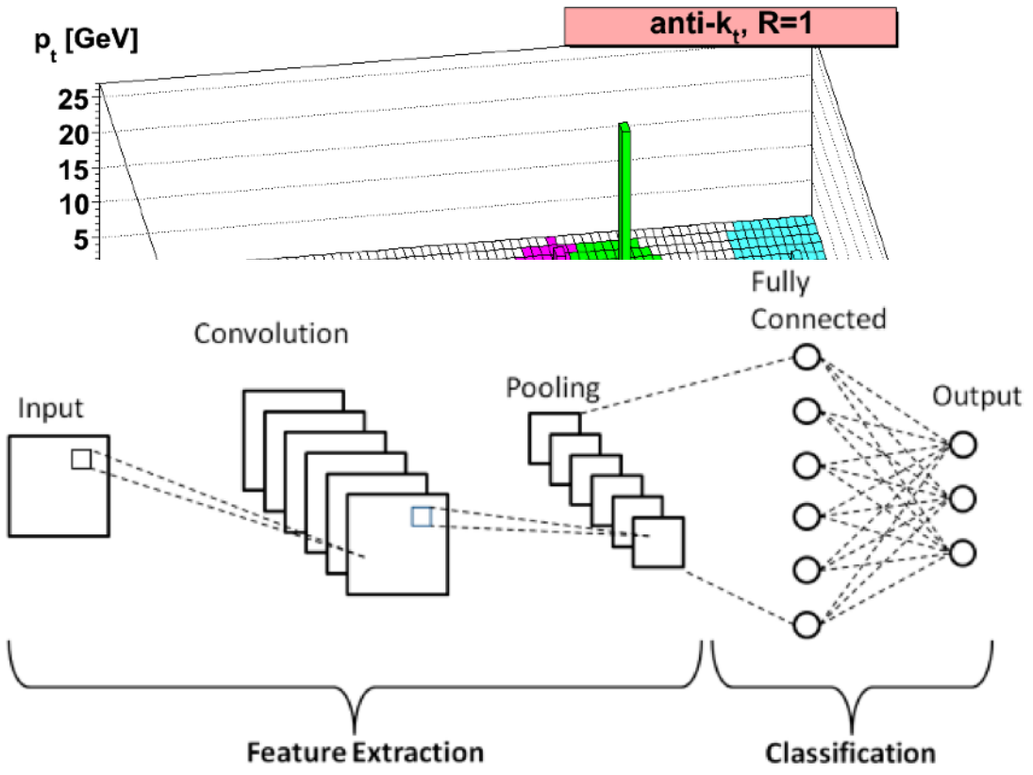
3. Novel ML techniques

   a. Clustering algorithm

   b. Binary classification w/ LDA

4. Performance comparison





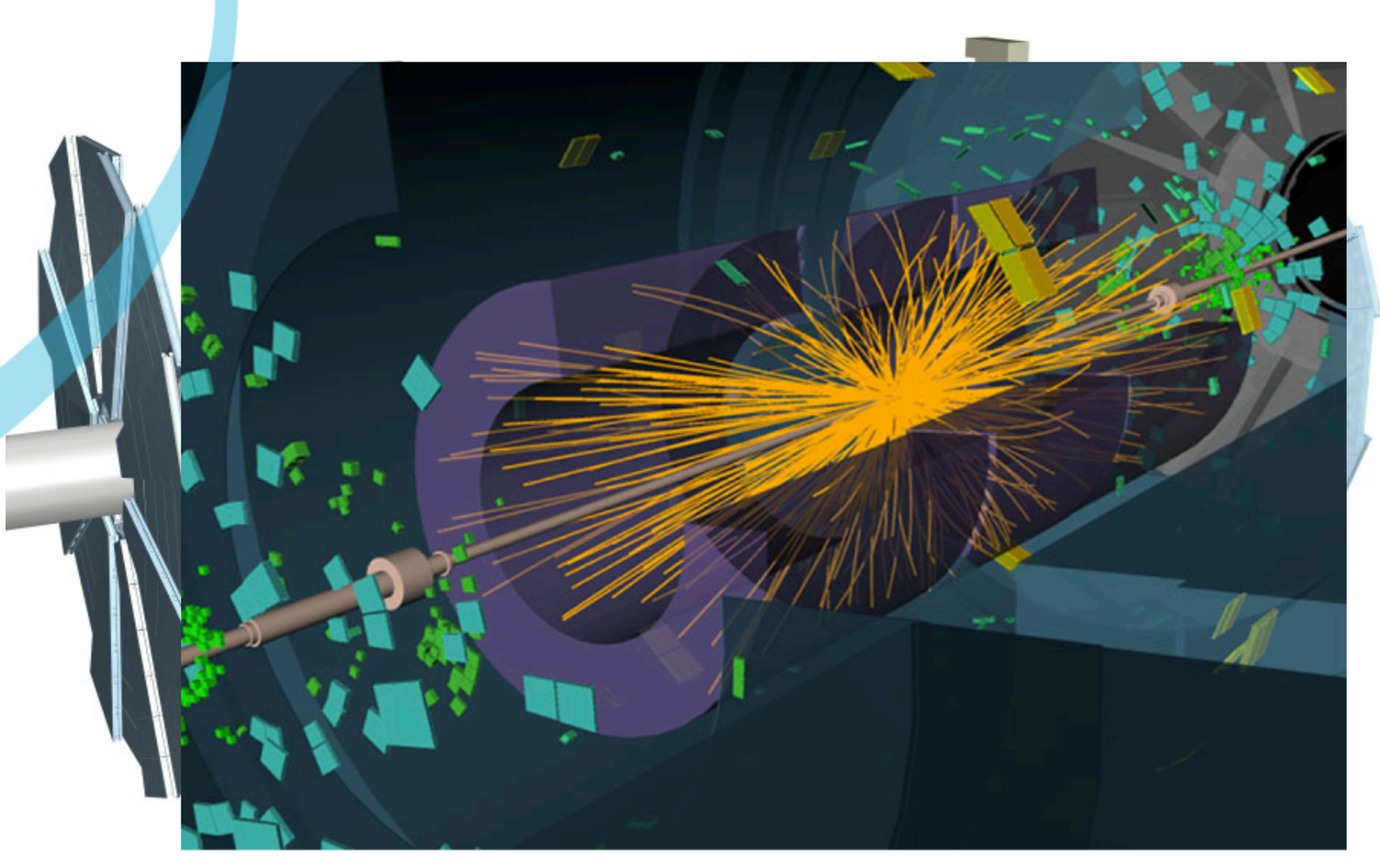
1. Data and its acquisition

1. Momentum of particles
2. Energy carried by particles
3. Capture position
4. Charge of particles
5. ...

features

1 event = 1 dataset with m examples

JETS = clusters in the data



1. Data and its acquisition

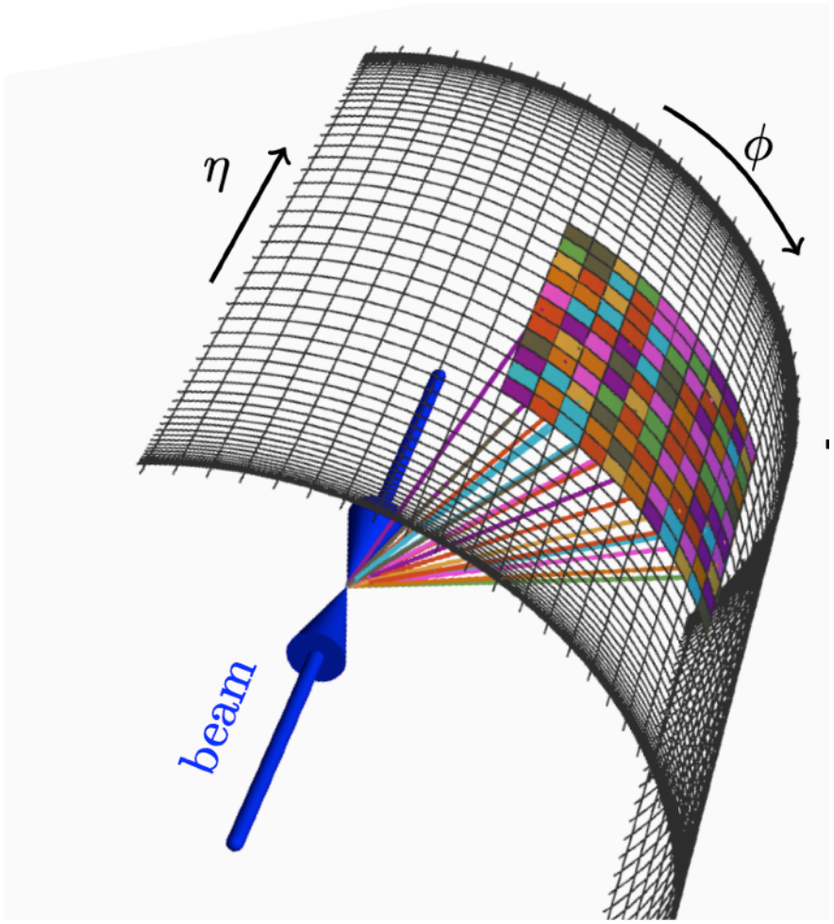


Image credits: CERN Doc Server

PF_dEta	PF_dPhi	PF_dR	PF_dTheta	PF_fromAK4Jet	PF_fromPV
[0.44430554, 0.43789667, 0.36538464, 0.3602575...	[-0.7161282, -0.3977437, -0.81739, -0.49003306...	[0.8427615, 0.5915687, 0.89533925, 0.6082088, ...	[-1.015492, -0.73738456, -1.1504285, -0.936854...	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, ...
[0.004705325, -0.004450232, 0.0057320073, -0....	[-0.0019089394, 0.043791108, 0.010397968, -0.0...	[0.0050778077, 0.04401665, 0.011873232, 0.0354...	[-0.38540852, 1.6720726, 2.074608, -2.0626178,...	[1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 2.0, 2.0, 3.0, 3.0, 3.0, 0.0, 0.0, 0.0, ...
[0.115596175, 0.0732975, 0.014152646, 0.011039...	[-0.010560029, -0.13039528, 0.0060405442, 0.01...	[0.11607752, 0.14958426, 0.015387838, 0.016447...	[-0.091099896, -1.0586973, 0.40340593, 0.83498...	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, ...
[0.06326231, 0.036711216, -0.273662, -0.384444...	[0.059476182, -0.08105969, 0.17061843, 0.02823...	[0.086830504, 0.08898531, 0.3224927, 0.3854794...	[0.7545608, -1.1455407, 2.5840986, 3.0682862, ...	[1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 2.0, 3.0, 3.0, 3.0, ...

2. The Machine-Learning problem

I: clustering

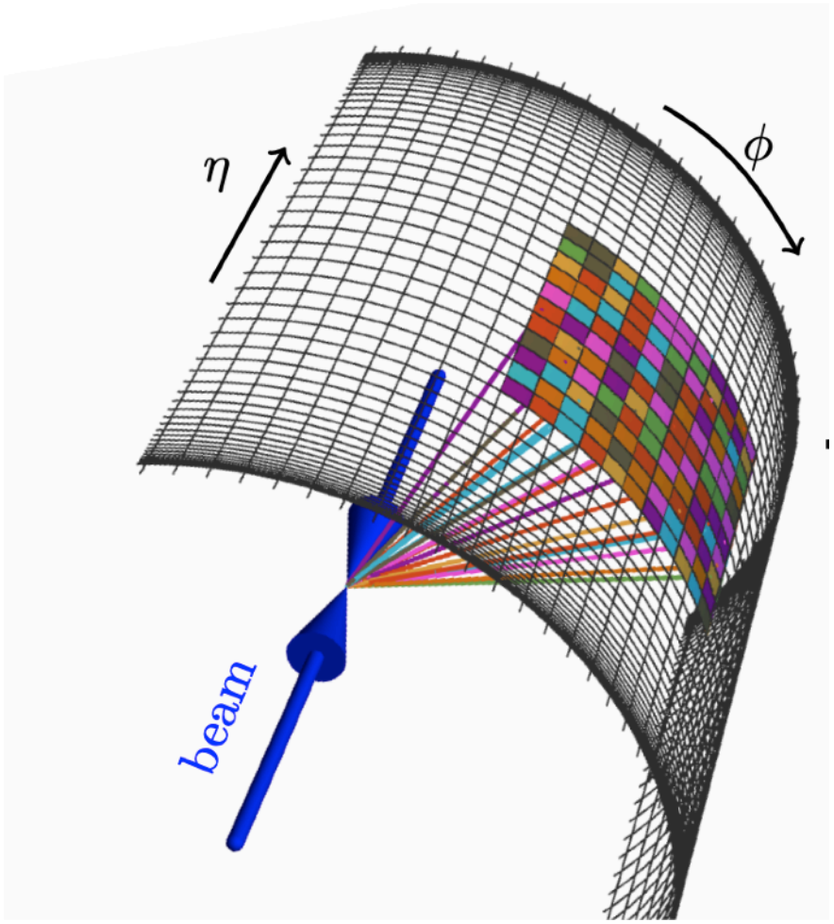


Image credits: CERN Doc Server

PF_dEta	PF_dPhi	PF_dR	PF_dTheta	PF_fromAK4Jet	PF_fromPV
[0.44430554, 0.43789667, 0.36538464, 0.3602575...	[-0.7161282, -0.3977437, -0.81739, -0.49003306...	[0.8427615, 0.5915687, 0.89533925, 0.6082088, ...	[-1.015492, -0.73738456, -1.1504285, -0.936854...	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, ...
[0.004705325, -0.004450232,).0057320073, -0....	[-0.0019089394, 0.043791108, 0.010397968, -0.0...	[0.0050778077, 0.04401665, 0.011873232, 0.0354...	[-0.38540852, 1.6720726, 2.074608, -2.0626178,...	[1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 2.0, 2.0, 3.0, 3.0, 3.0, 0.0, 0.0, 0.0, ...
[0.115596175, 0.0732975, 0.014152646, 0.011039...	[-0.010560029, -0.13039528, 0.0060405442, 0.01...	[0.11607752, 0.14958426, 0.015387838, 0.016447...	[-0.091099896, -1.0586973, 0.40340593, 0.83498...	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, ...
[0.06326231, 0.036711216, -0.273662, -0.384444...	[0.059476182, -0.08105969, 0.17061843, 0.02823...	[0.086830504, 0.08898531, 0.3224927, 0.3854794...	[0.7545608, -1.1455407, 2.5840986, 3.0682862, ...	[1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 2.0, 3.0, 3.0, 3.0, ...

2. The Machine-Learning problem

II: classification

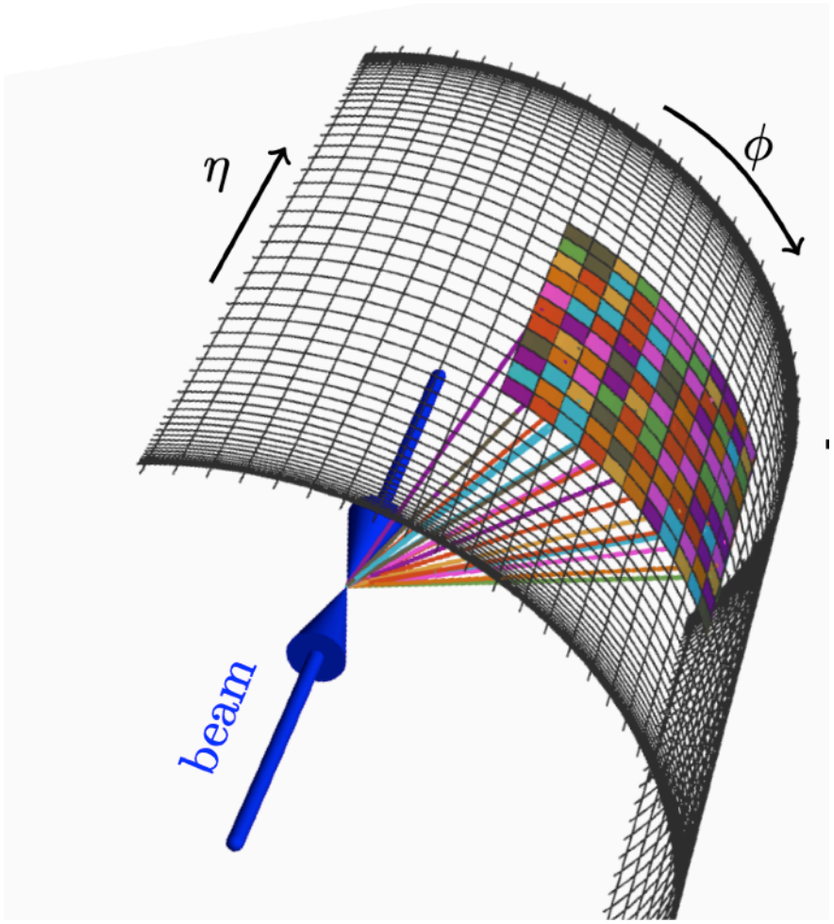
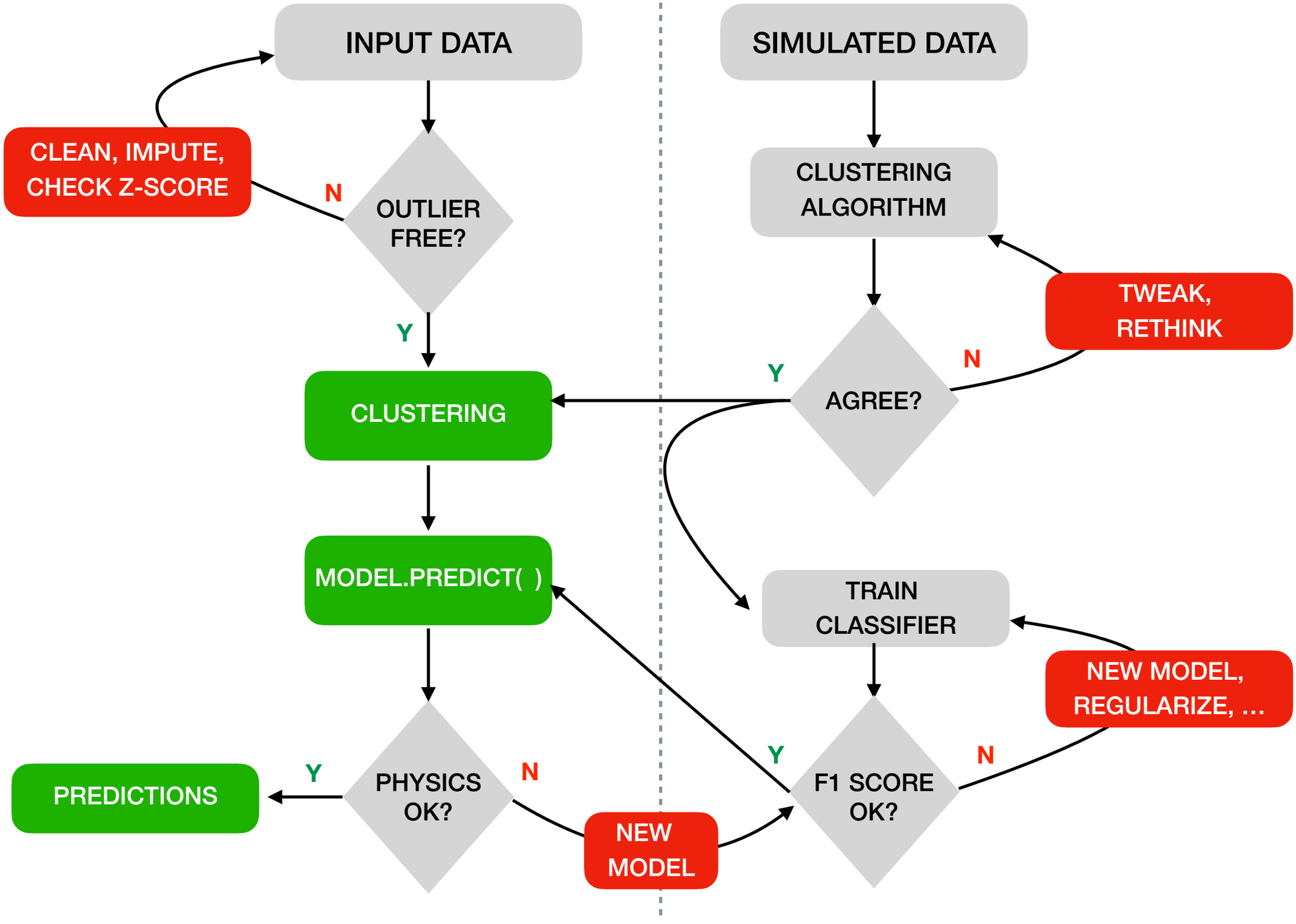


Image credits: CERN Doc Server

	PF_dEta	PF_dPhi	PF_dR	PF_dTheta	PF_fromAK4Jet	PF_fromPV
?	[0.44430554, 0.43789667, 0.36538464, 0.3602575...	[-0.7161282, -0.3977437, -0.81739, -0.49003306...	[0.8427615, 0.5915687, 0.89533925, 0.6082088, ...	[-1.015492, -0.73738456, -1.1504285, -0.936854...	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, ...
?	[0.004705325, -0.004450232,).0057320073, -0....	[-0.0019089394, 0.043791108, 0.010397968, -0.0...	[0.0050778077, 0.04401665, 0.011873232, 0.0354...	[-0.38540852, 1.6720726, 2.074608, -2.0626178,...	[1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 2.0, 2.0, 3.0, 3.0, 3.0, 0.0, 0.0, 0.0, ...
?	[0.115596175, 0.0732975, 0.014152646, 0.011039...	[-0.010560029, -0.13039528, 0.0060405442, 0.01...	[0.11607752, 0.14958426, 0.015387838, 0.016447...	[-0.091099896, -1.0586973, 0.40340593, 0.83498...	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, ...
?	[0.06326231, 0.036711216, -0.273662, -0.384444...	[0.059476182, -0.08105969, 0.17061843, 0.02823...	[0.086830504, 0.08898531, 0.3224927, 0.3854794...	[0.7545608, -1.1455407, 2.5840986, 3.0682862, ...	[1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 2.0, 3.0, 3.0, 3.0, ...

? = A or B

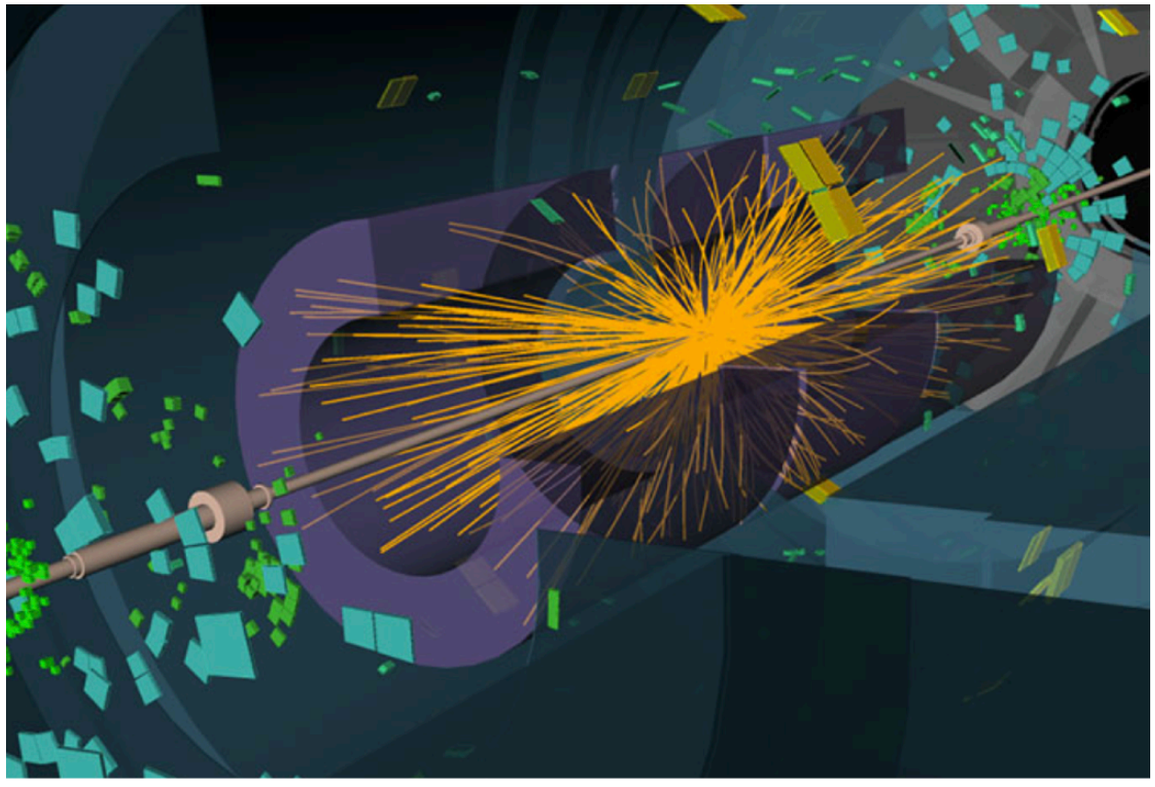
2. The Machine-Learning problem

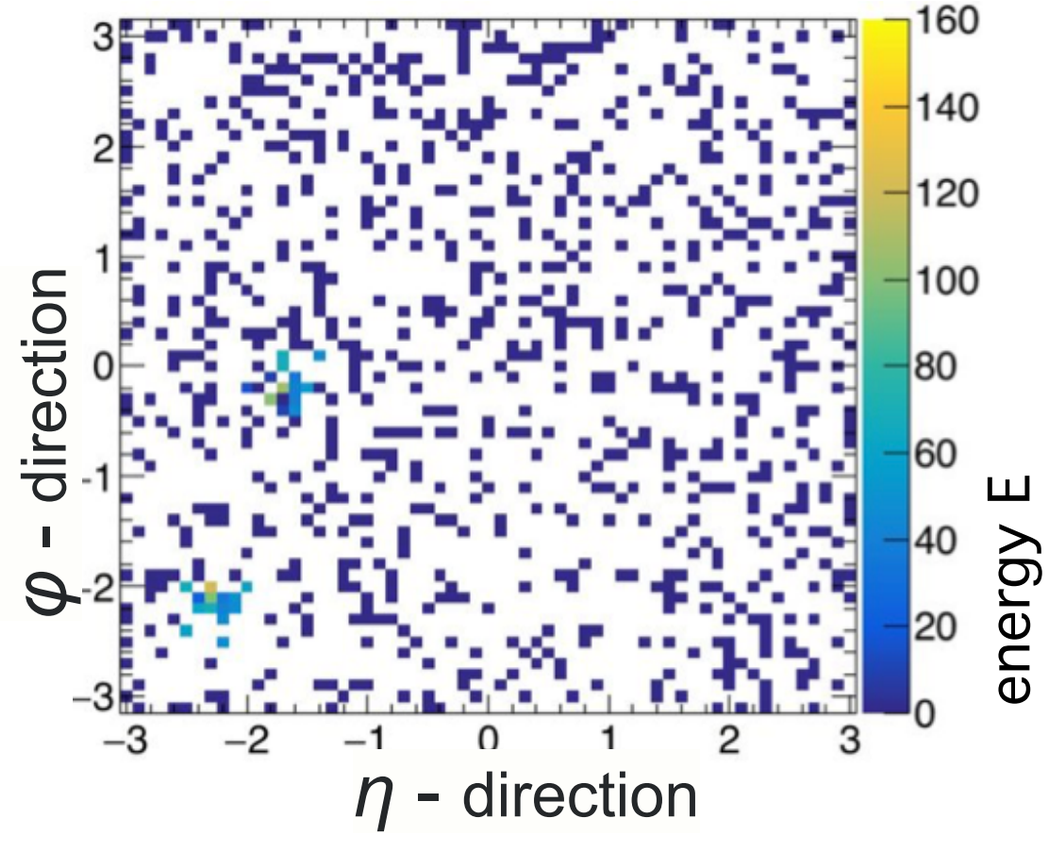


3.a) The clustering algorithm

1 event = 1 dataset with m examples features = $(\eta, \varphi, E, \dots)$

JETS = clusters in the data





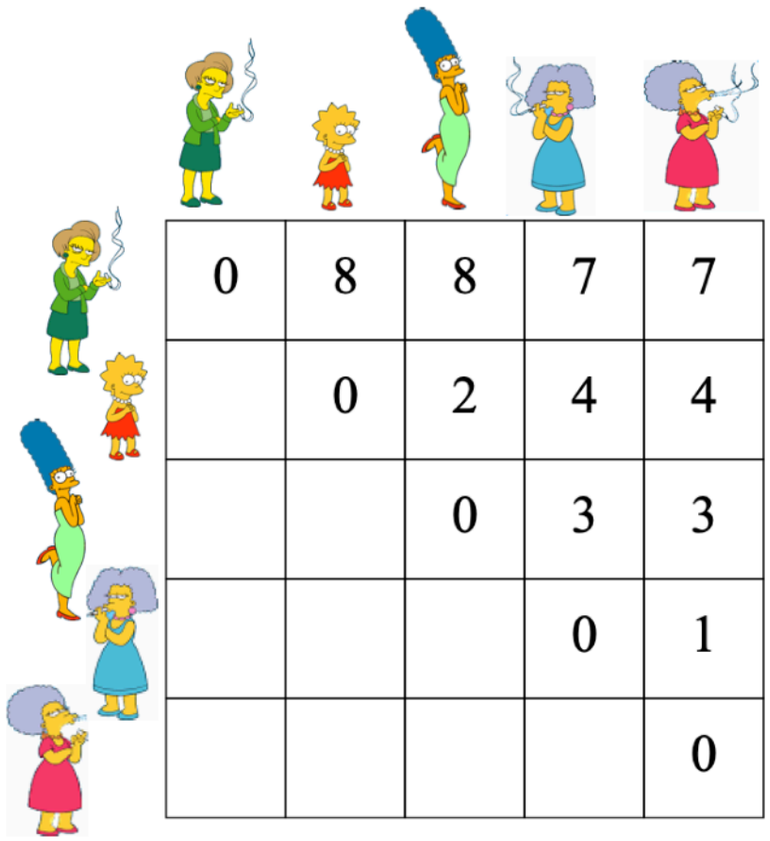
Physical intuition: clusters formed around datapoint with high E.

What is the loss function for our clustering algorithm?

3.a) The clustering algorithm

1 event = 1 dataset with m examples

features = $(\eta, \varphi, E, \dots)$

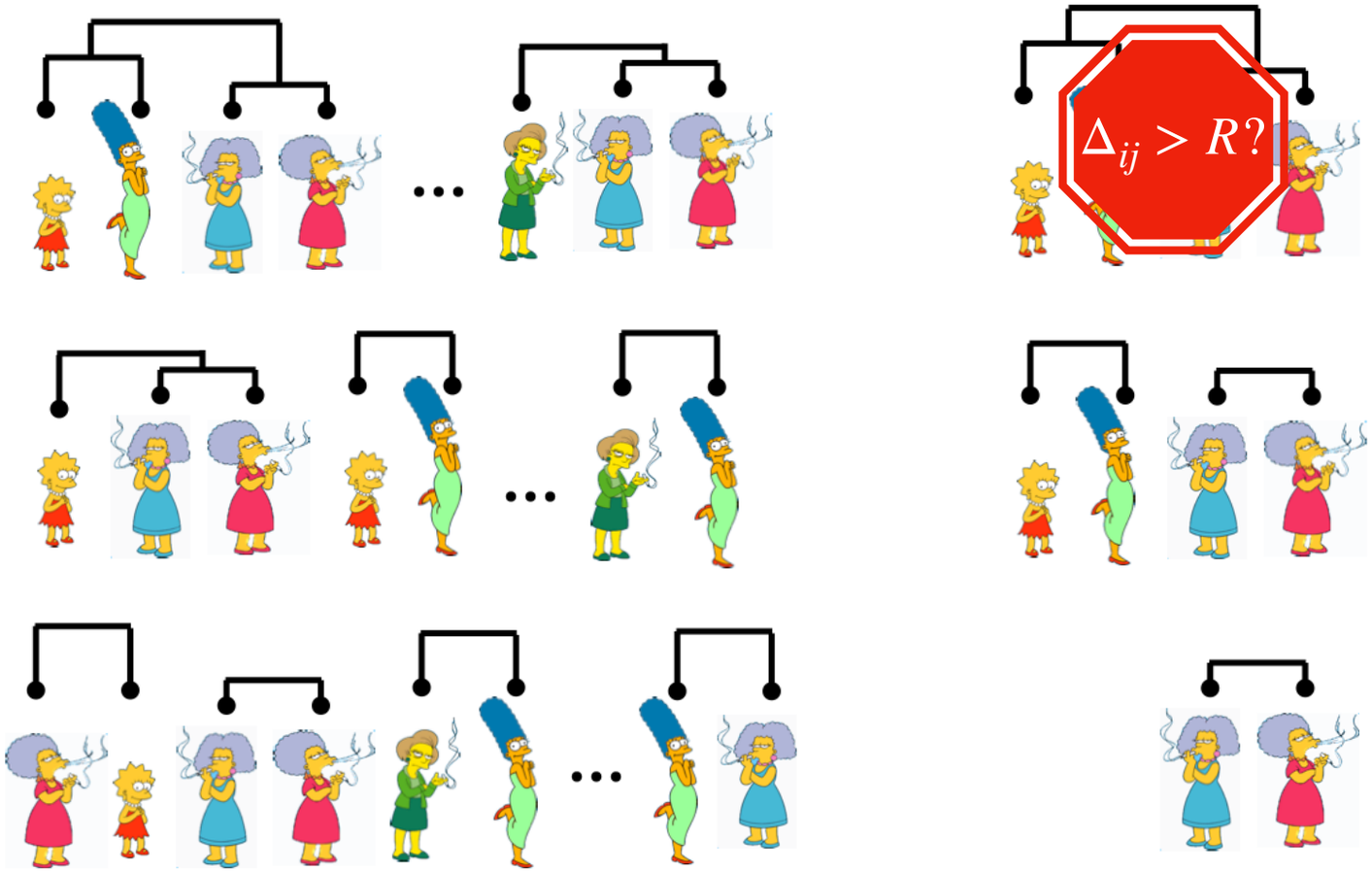


0	8	8	7	7
	0	2	4	4
		0	3	3
			0	1
				0

m x m distance matrix d_{ij}

$$d_{ij} = \min(E_i^p, E_j^p)\frac{\Delta_{ij}}{R}$$

$$\Delta_{ij} = (\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2$$



all possible merges

choose the min

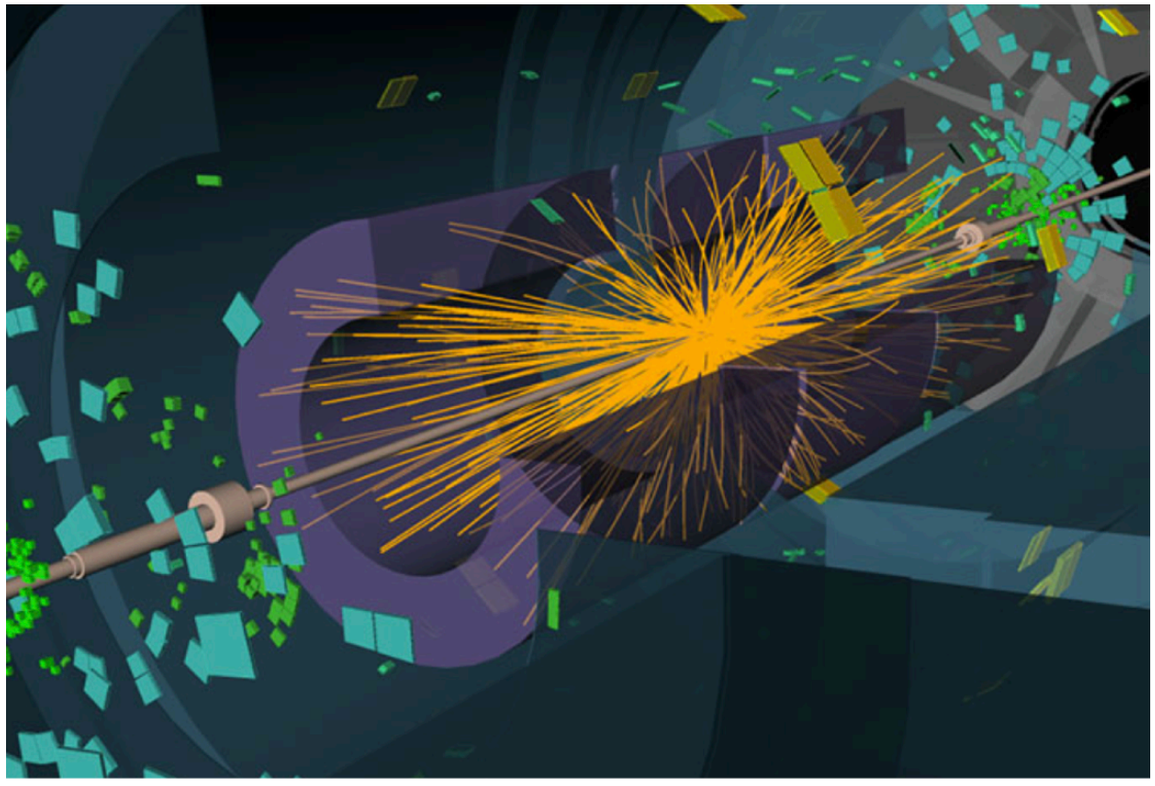
only merge if $\Delta_{ij} \leq R$

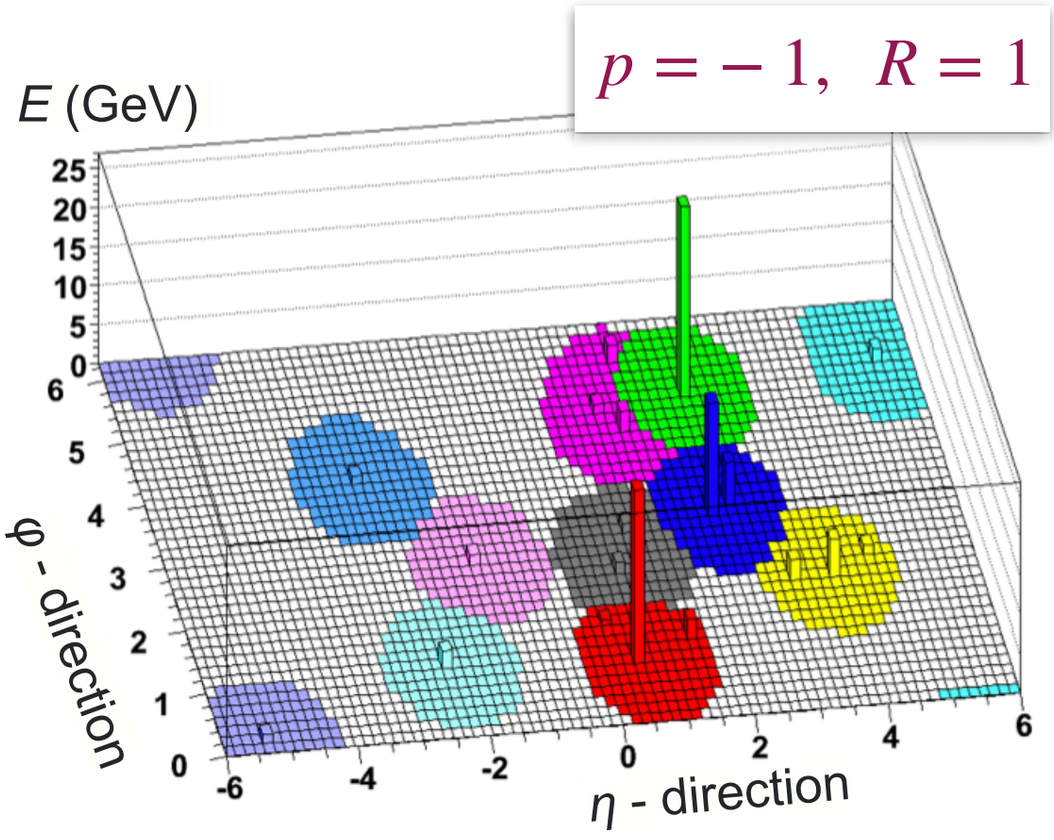
PRL **130**, 162001 **S. Dutta**, G. Salam, et al.

3.a) The clustering algorithm

1 event = 1 dataset with m examples features = $(\eta, \varphi, E, \dots)$

JETS = clusters in the data





Physical intuition: clusters formed around datapoint with highest E.

Questions: What is the loss function? Which clustering algorithm?

PRL **130**, 162001 **S. Dutta**, G. Salam, et al.

Overview of this talk

✓1. Data and its acquisition

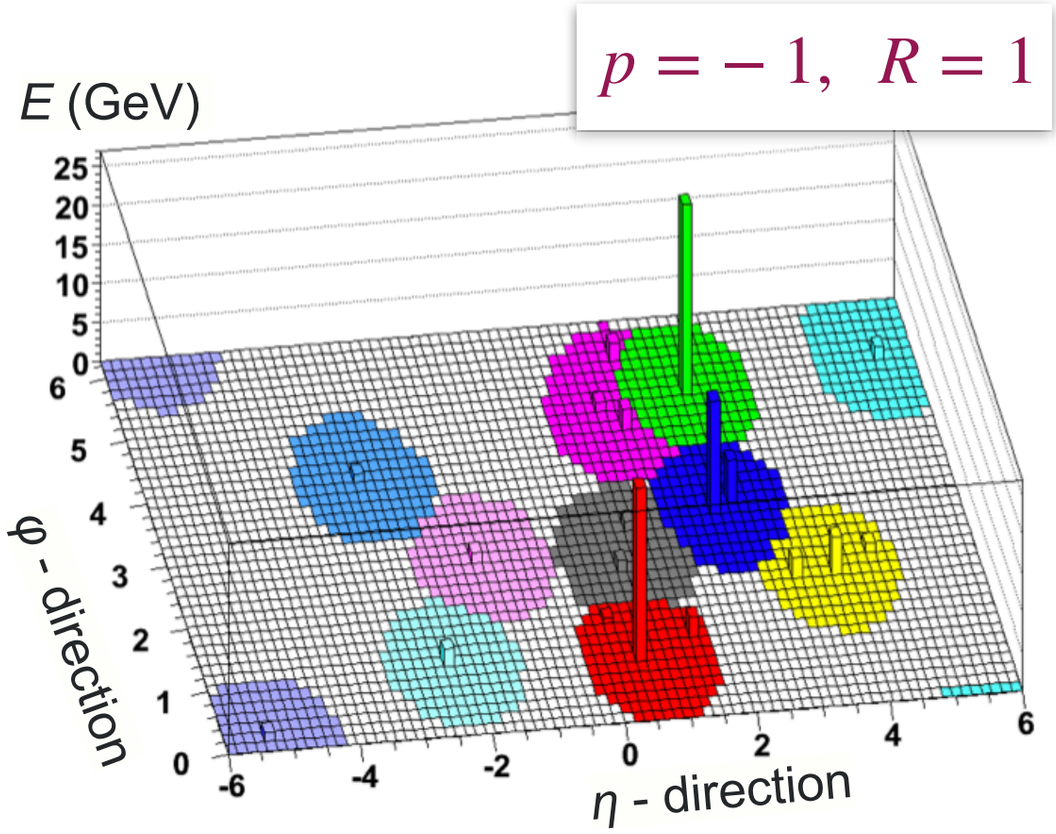
✓2. The ML problem statement

3. Novel ML techniques

✓a. Clustering algorithm

b. Binary classification w/ LDA

4. Performance comparison

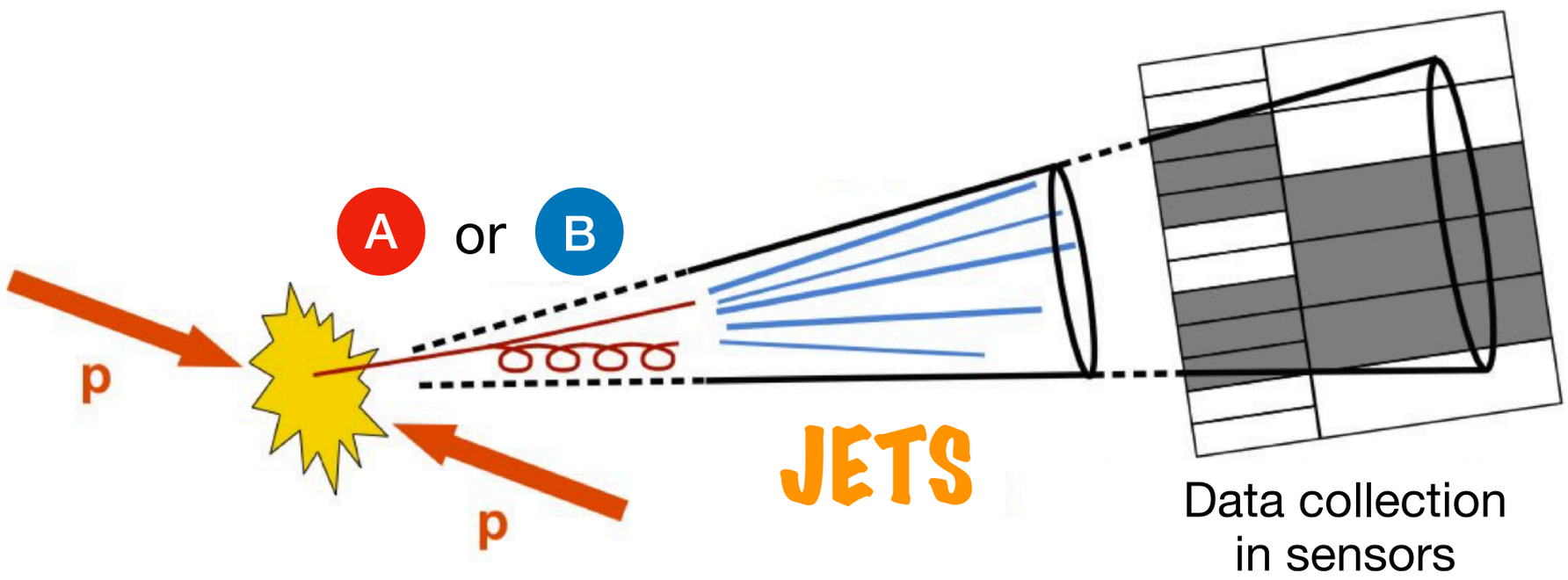


3.b) The classification problem

What are we classifying?

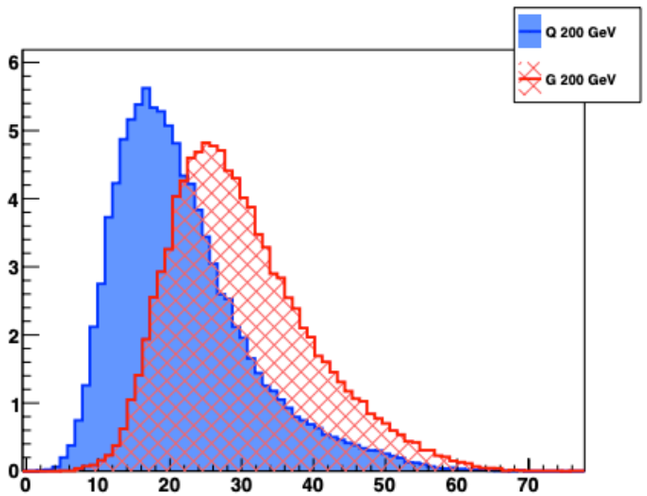
? = A or B

	PF_dEta	PF_dPhi	PF_dR	PF_dTheta	PF_fromAK4Jet	PF_fromPV
A	[0.44430554, 0.43789667, 0.36538464, 0.3602575...	[-0.7161282, -0.3977437, -0.81739, -0.49003306...	[0.8427615, 0.5915687, 0.89533925, 0.6082088, ...	[-1.015492, -0.73738456, -1.1504285, -0.936854...	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, ...
B	[0.004705325, -0.004450232,).0057320073, -0....	[-0.0019089394, 0.043791108, 0.010397968, -0.0...	[0.0050778077, 0.04401665, 0.011873232, 0.0354...	[-0.38540852, 1.6720726, 2.074608, -2.0626178,...	[1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 2.0, 2.0, 3.0, 3.0, 3.0, 0.0, 0.0, 0.0, ...

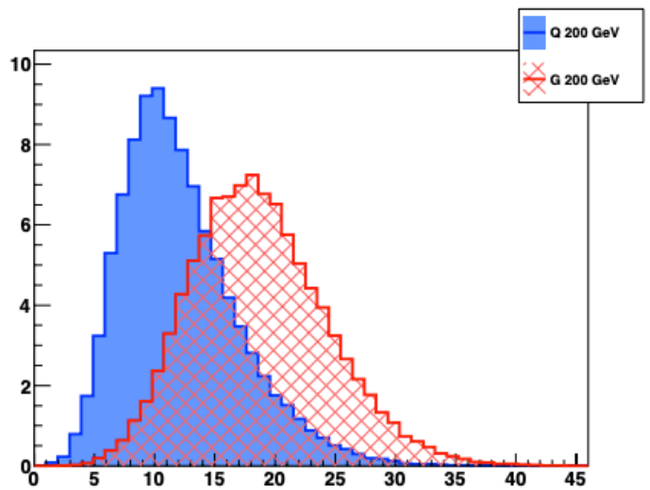


Digression: EDA

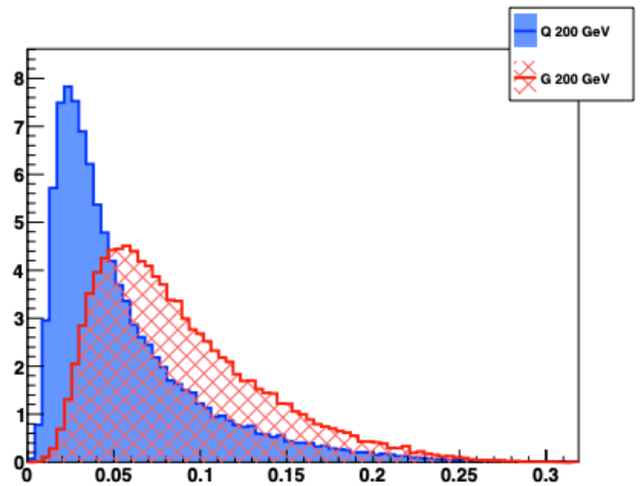
Jet mass



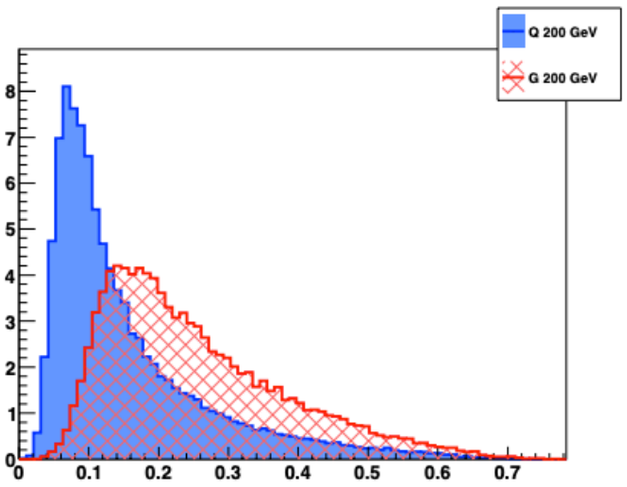
Charged particle count



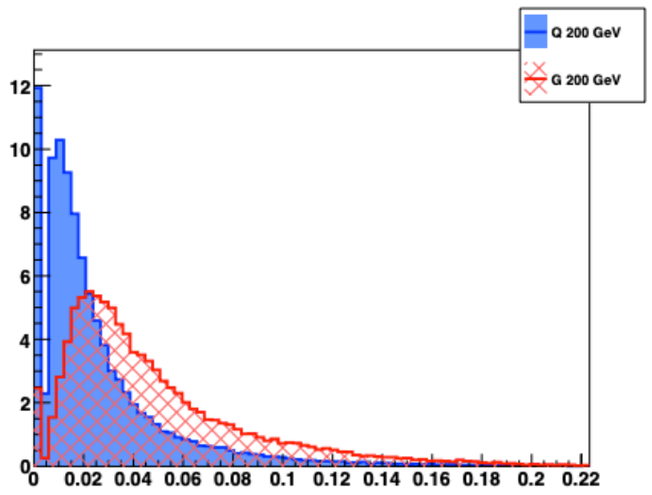
Jet broadening



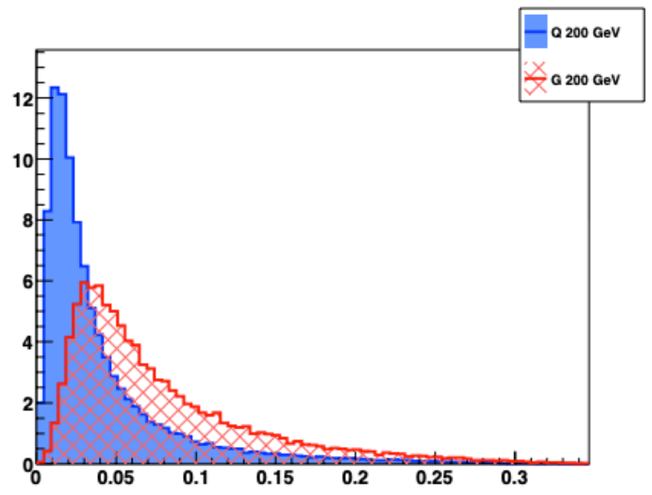
Jet angularity



P_T fraction of 3rd hardest subjet



Moment of Hu

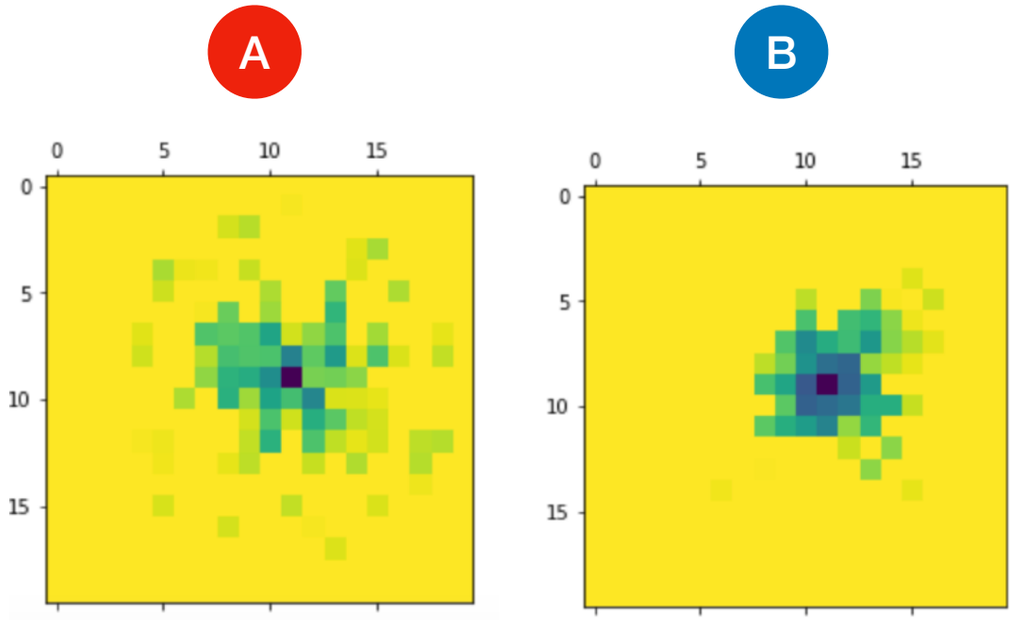


Best F1-score (w/Decision Trees): 0.68

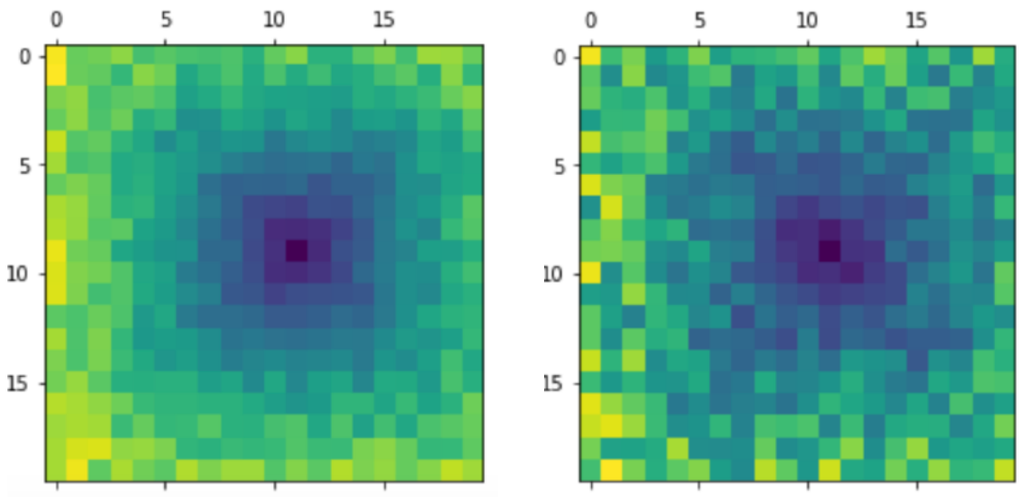
A B

[For more exploratory analysis: https://github.com/TheKivs/LHC_Jet_Tagging]

3.b) The classification problem



PF_dR, single clusters

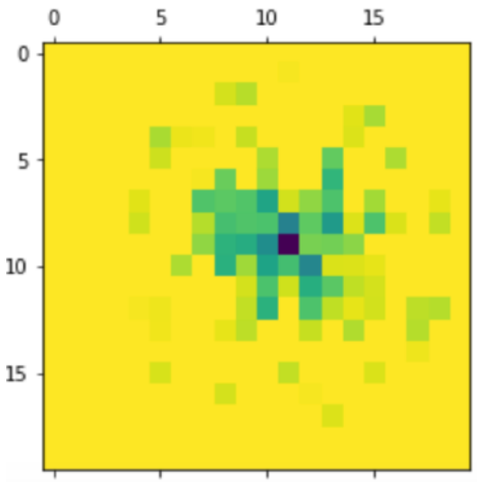


PF_dR, averaged clusters

	η	φ	PF_dR	PF_dTheta	PF_fromAK4Jet	PF_fromPV
A	[0.44430554, 0.43789667, 0.36538464, 0.3602575...	[-0.7161282, -0.3977437, -0.81739, -0.49003306...	[0.8427615, 0.5915687, 0.89533925, 0.6082088, ...	[-1.015492, -0.73738456, -1.1504285, -0.936854...	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, ...
B	[0.004705325, -0.004450232, .0057320073, -0....	[-0.0019089394, 0.043791108, 0.010397968, -0.0...	[0.0050778077, 0.04401665, 0.011873232, 0.0354...	[-0.38540852, 1.6720726, 2.074608, -2.0626178,...	[1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 2.0, 2.0, 3.0, 3.0, 3.0, 0.0, 0.0, 0.0, ...
B	[0.115596175, 0.0732975, 0.014152646, 0.011039...	[-0.010560029, -0.13039528, 0.0060405442, 0.01...	[0.11607752, 0.14958426, 0.015387838, 0.016447...	[-0.091099896, -1.0586973, 0.40340593, 0.83498...	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.0, ...
A	[0.06326231, 0.036711216, -0.273662, -0.384444...	[0.059476182, -0.08105969, 0.17061843, 0.02823...	[0.086830504, 0.08898531, 0.3224927, 0.3854794...	[0.7545608, -1.1455407, 2.5840986, 3.0682862, ...	[1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	[3.0, 3.0, 2.0, 3.0, 3.0, 2.0, 3.0, 3.0, 3.0, ...

Idea: For each cluster, treat features as the color intensity of a 2D picture

3.b) The classification problem



$= \overrightarrow{X}$ k^2 - dimensional vector

Aim: find vector $\overrightarrow{v}$ such that

$$\textbf{argmax}_v \; J(\overrightarrow{v}) = \frac{(\hat{\mu}_A - \hat{\mu}_B)^2}{\hat{s}_A^2 + \hat{s}_B^2}$$

$$\hat{\mu}_A = \frac{1}{|n_A|} \sum_{C_i=A} \overrightarrow{X}_{(i)} \cdot \overrightarrow{v}$$

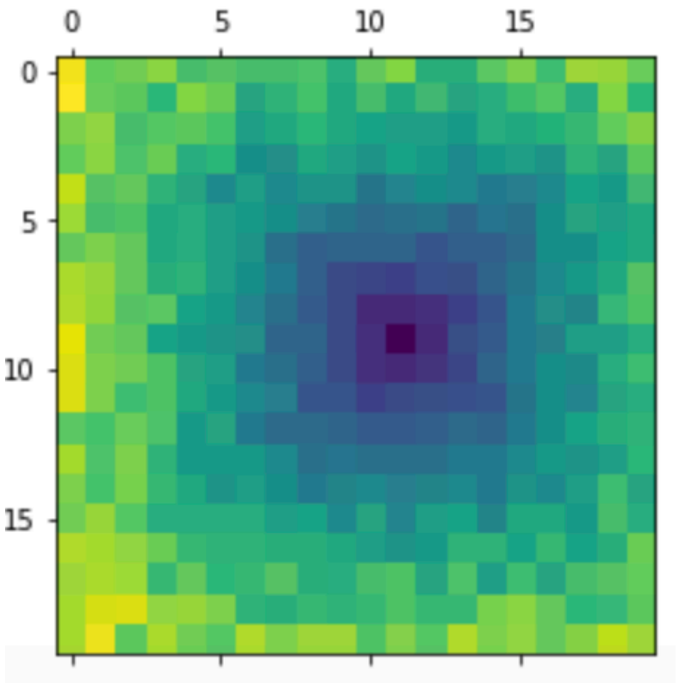
FISHER'S LINEAR DISCRIMINANT

$$\mathscr{D}(X) = \overrightarrow{X} \cdot \overrightarrow{v}$$

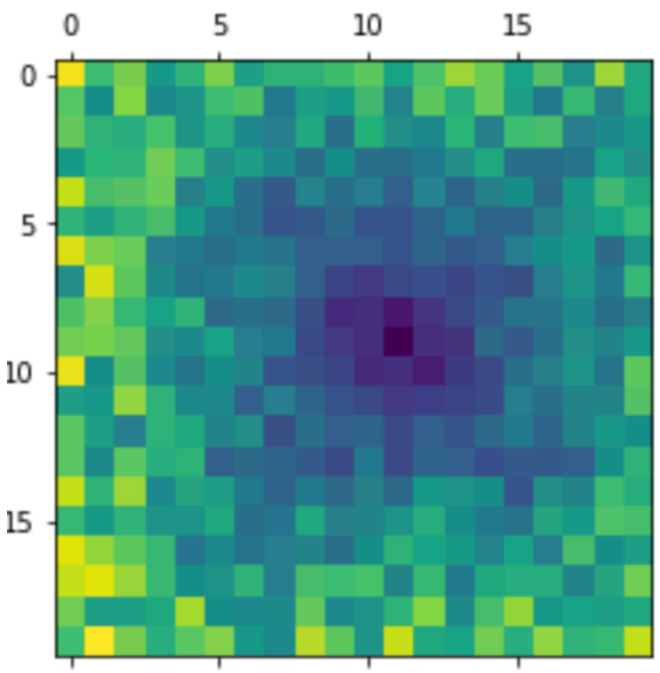
$\mathscr{D}[X] < 0$ A

$\mathscr{D}[X] > 0$ B

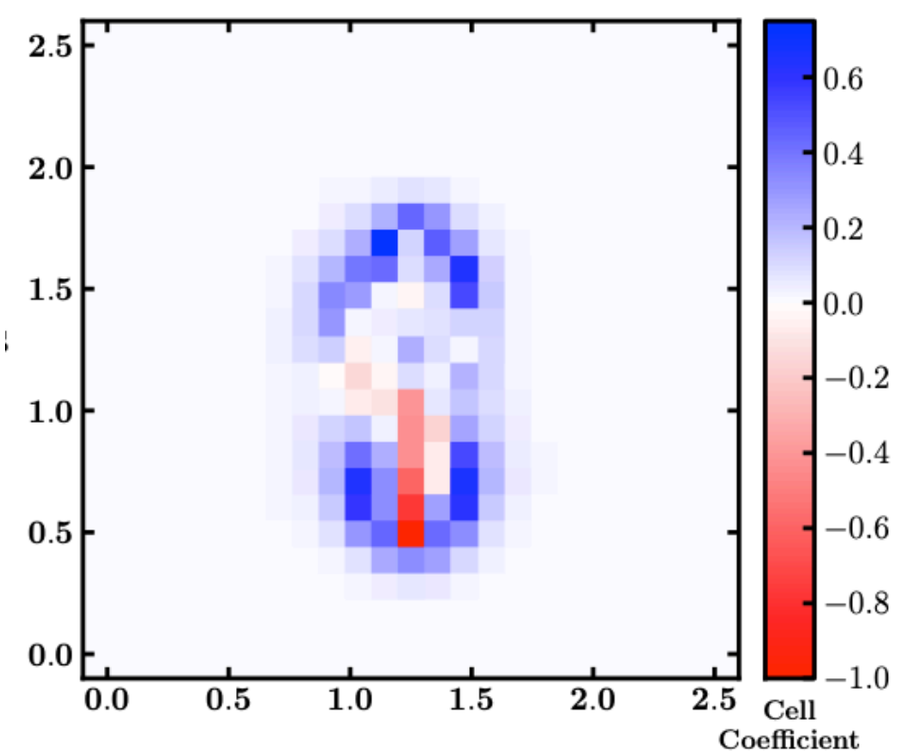
3.b) The classification problem



A



B



Fisher jet image $\overrightarrow{v}$

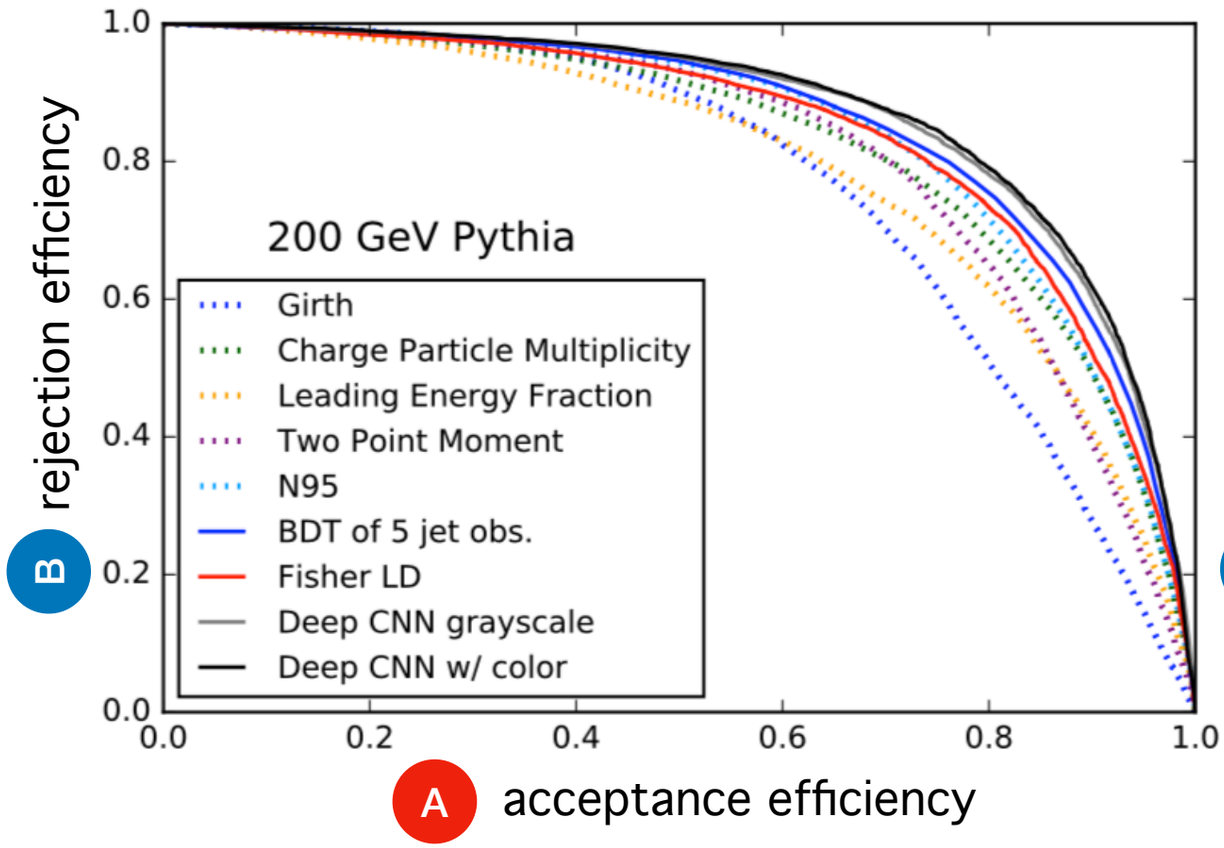
FISHER'S LINEAR DISCRIMINANT

$$\mathscr{D}(X) = \overrightarrow{X} \cdot \overrightarrow{v}$$

$\mathscr{D}[X] < 0$ → A

$\mathscr{D}[X] > 0$ → B

4. Performance comparison



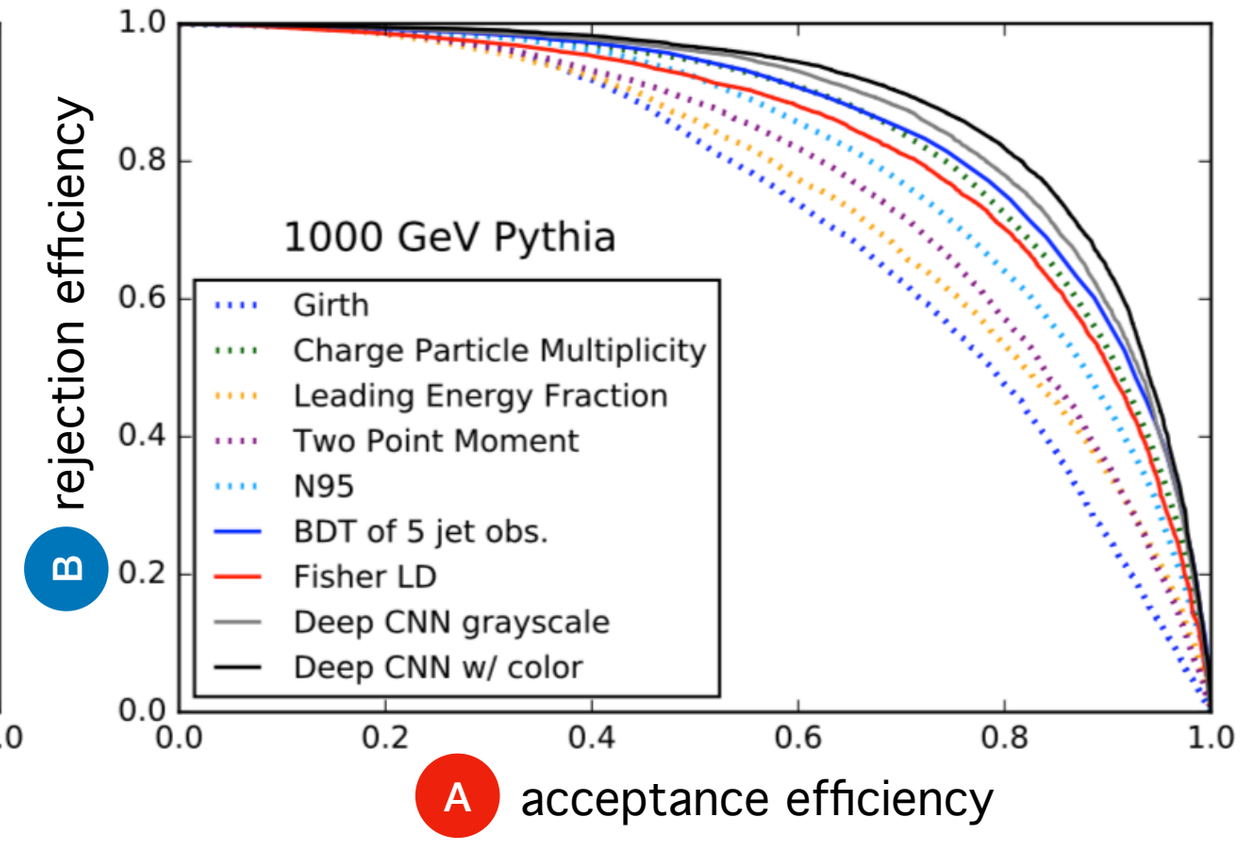


Image credits: Matt Schwartz, Harvard

- ROC: highest attainable B rejection efficiency at given A acceptance efficiency
- FLD performs significantly better than ML using physics-motivated variables
- AUC for FLD = 0.77, AUC for NN = 0.84