

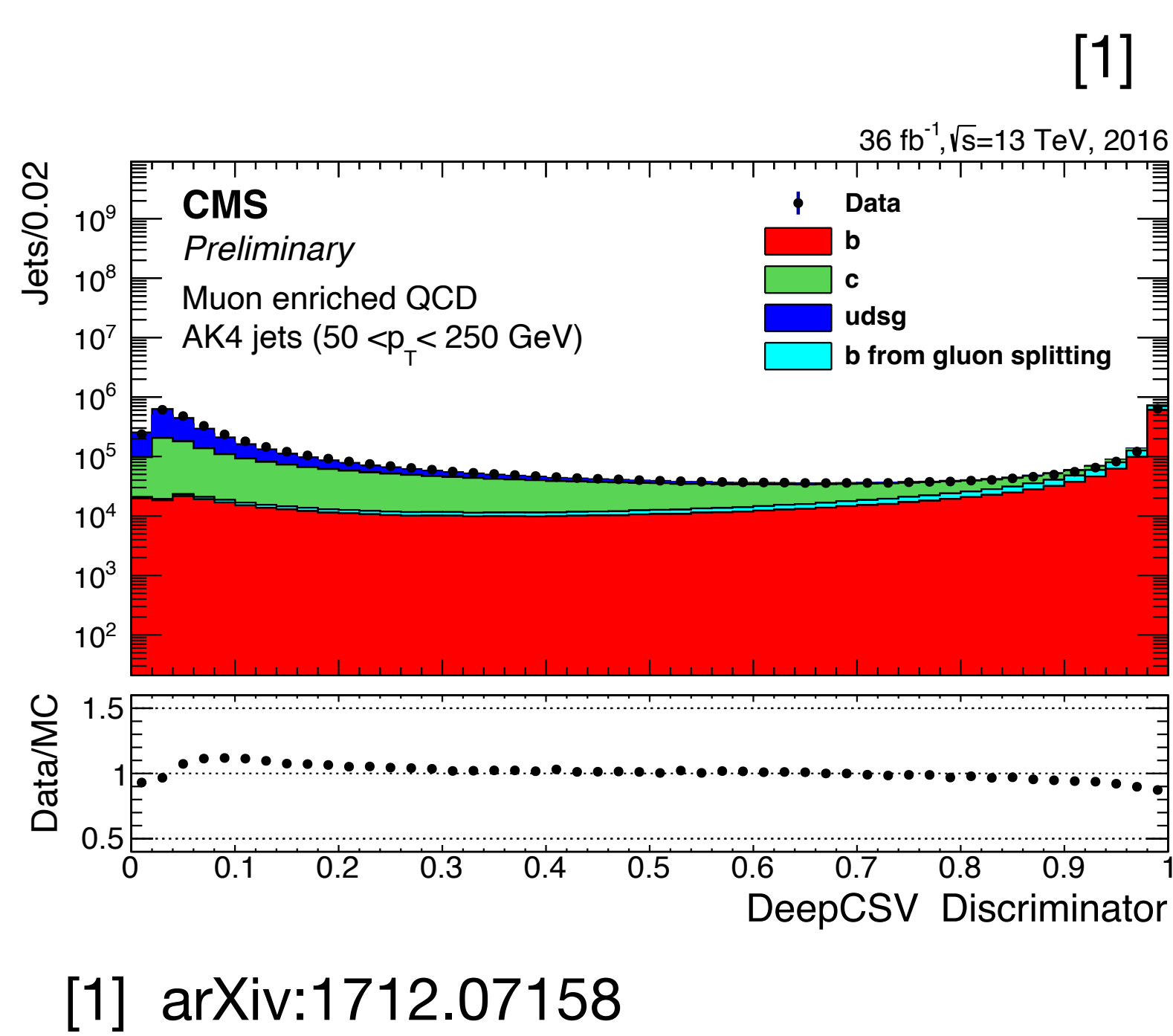
Adversarial Neural Network based shape calibrations of observables for jet-tagging at CMS

Martin Erdmann, Benjamin Fischer, Jan Middendorf, Dennis Noll, Yannick Rath, Marcel Rieger, Erwin Rudi, David Schmidt

ACAT 2021 - 29 Nov to 3 Dec
Contact: erwin.rudi@rwth-aachen.de

Heavy-flavor jet tagging

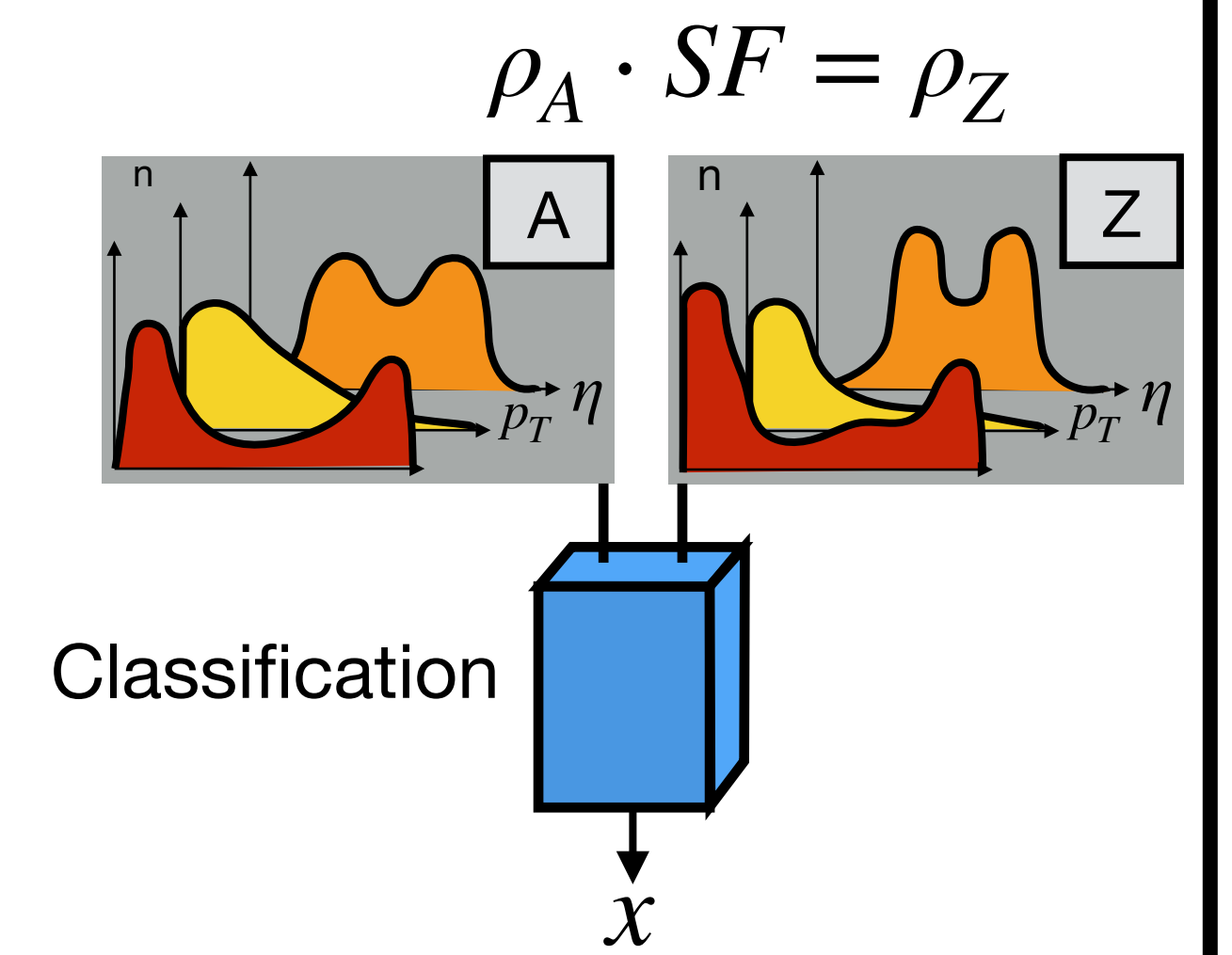
- DL algorithms discriminate between flavors
- Assign flavor discriminator score to jet: $P(B), P(C), P(L)$
- Data-Simulation differences observed
→ Need for calibration



Method

- Train classifier between 2 classes
- Use BCE loss
- Classifier measures ratio
- Output gives continuous SFs

$$SF = e^x = \frac{\rho_Z}{\rho_A}$$



- Subtraction can be written as ratio $w_{A \rightarrow Z}(x) = (1 - SF(x)) \cdot \frac{N_Z^{tot}}{N_A^{tot}}$

Preprocessing on event level

Data Selection

Light enriched Bottom enriched Charm: signal + 50% bkg Charm: 50% bkg

Legend

Data: ■ Simulation: ■

6 variables: ■ Kin: p_T, η
tag: $P(B), P(C)$
 C_{vsB}, C_{vsL}

3 flavors: ■ bottom
■ charm
■ light

reweighting: \otimes

- 1 **Select 4 regions**
 - Bottom enriched and Light enriched
 - 2 charm regions with background contamination
- 2 **Create charm enriched region**
 - Remove background by subtraction
 - Classify between the two charm regions
 - Reweight one region to look like the subtraction
 - Event structure is preserved
- 3 **Prevent changes of kin. variables in main calibration**
 - Classify between kin. variables for all regions
 - Reweight Simulation events

Result

- Reweighted distribution looks like signal
- More charm purity

Result: Kin. variables of Simulation look like Data

Main Calibration on jet level

- 4 **Adversarial network produces SF for every jet**

$L_{tot} = \text{BCE}(X_D, \text{label}) + a \cdot \text{MSE}(X_G, X_{ideal})$
 $X_{ideal} = X_G + X_D$
- 4 **Correct Data-Simulation differences for tagging variables**
 - Generator produces SFs for every jet depending on flavor
 - Discriminator measures ratio between Data and scaled Simulation
 - Generator receives feedback in loss from Discriminator
- 5 **Result of main calibration**
 - Continuous SFs for bottom, charm and light flavor
 - Results comparable to traditional methods
 - Kin. variables are not changed significantly
 - Successful calibration of tagging variables in light-, charm- and bottom enriched regions

Application

Final goal

- Tagging variables look like data
- Kinematic variables stay the same

[2] <https://cds.cern.ch/record/266647?ln=en>