中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

CMS

# CMS jet tagging

**Zhenxuan Zhang 1**

Working group meeting

1 IHEP of Chinese Academy of Sciences (CN)

# outline

contents

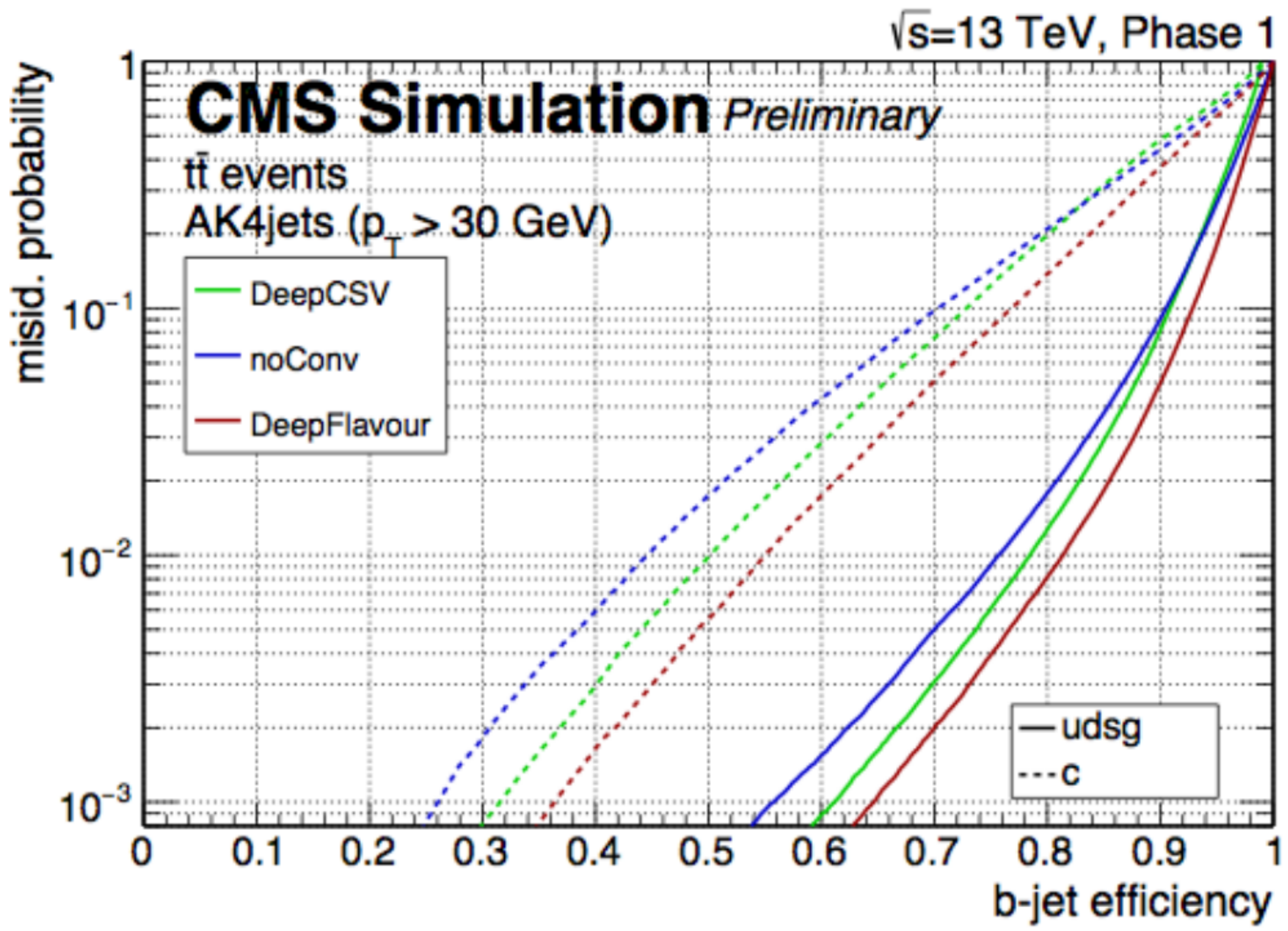
- Introduction
- AK4 jets tagging
  - DeepCSV model
  - DeepFlavour model
- AK8 jets tagging
  - DeepAK8 model
  - ParticleNet model
  - ParticleTransformer model
- Calibration of the taggers
  - SFsBDT method
  - LundJetPlane method

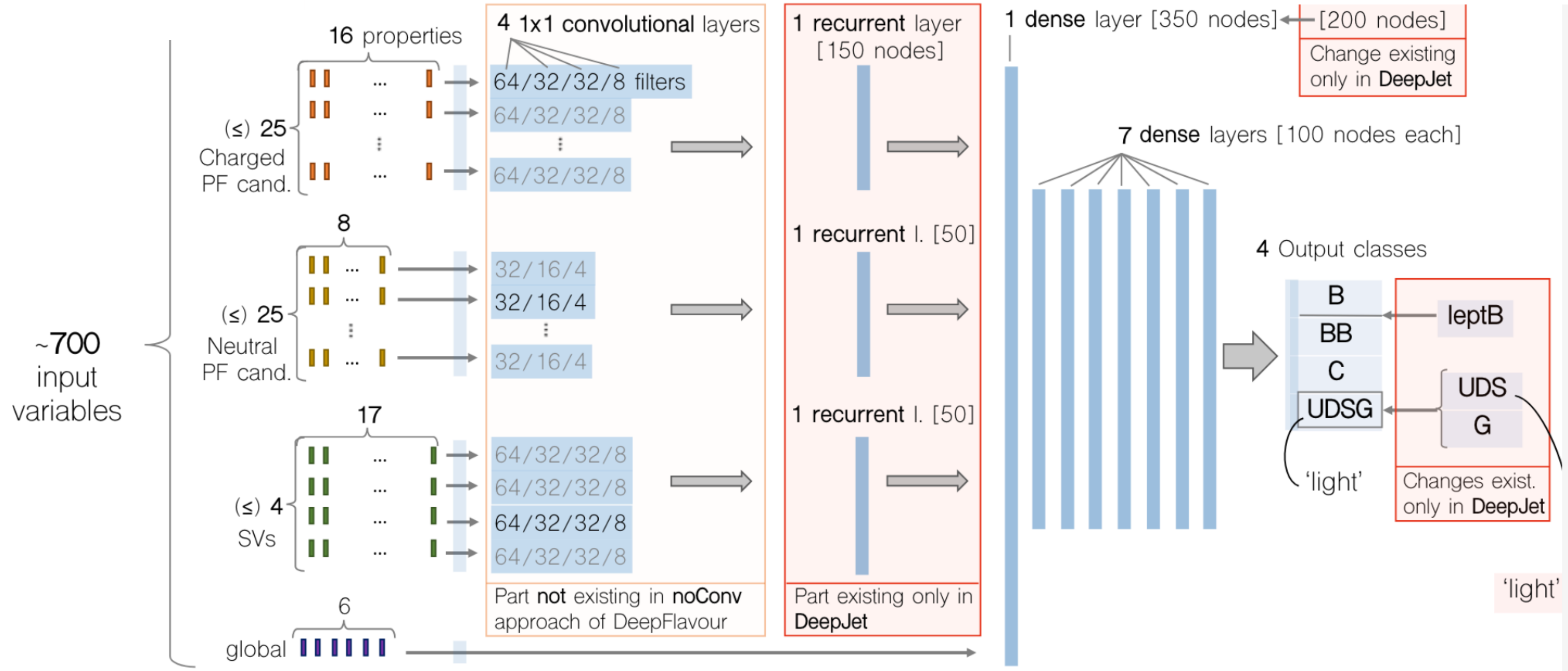
# Introduction

- Jet is one of the most ubiquitous but also fascinating objects at the LHC
- Very common in CMS to use ML method to tag jets:
  - heavy flavour tagging (bottom/charm)
  - heavy resonance tagging (top/W/Z/Higgs)
  - quark/gluon discrimination
  - ...
- One of the most active areas for ML
  - lots of deep approaches have been proposed in the past few years
  - trending: low-level inputs + deep learning

# AK4 jets tagging

- DeepCSV model(CMS DP 2017-005 cds.cern.ch/record/2255736) :
  - Pure deep NN, multi-classification with combine secondary vertex inputs
- ★ DeepFlavour model (CMS DP 2017-013 cds.cern.ch/record/2263802)
  - Larger inputs, deeper NN + convolutional, recurrent layers
  - Best so far



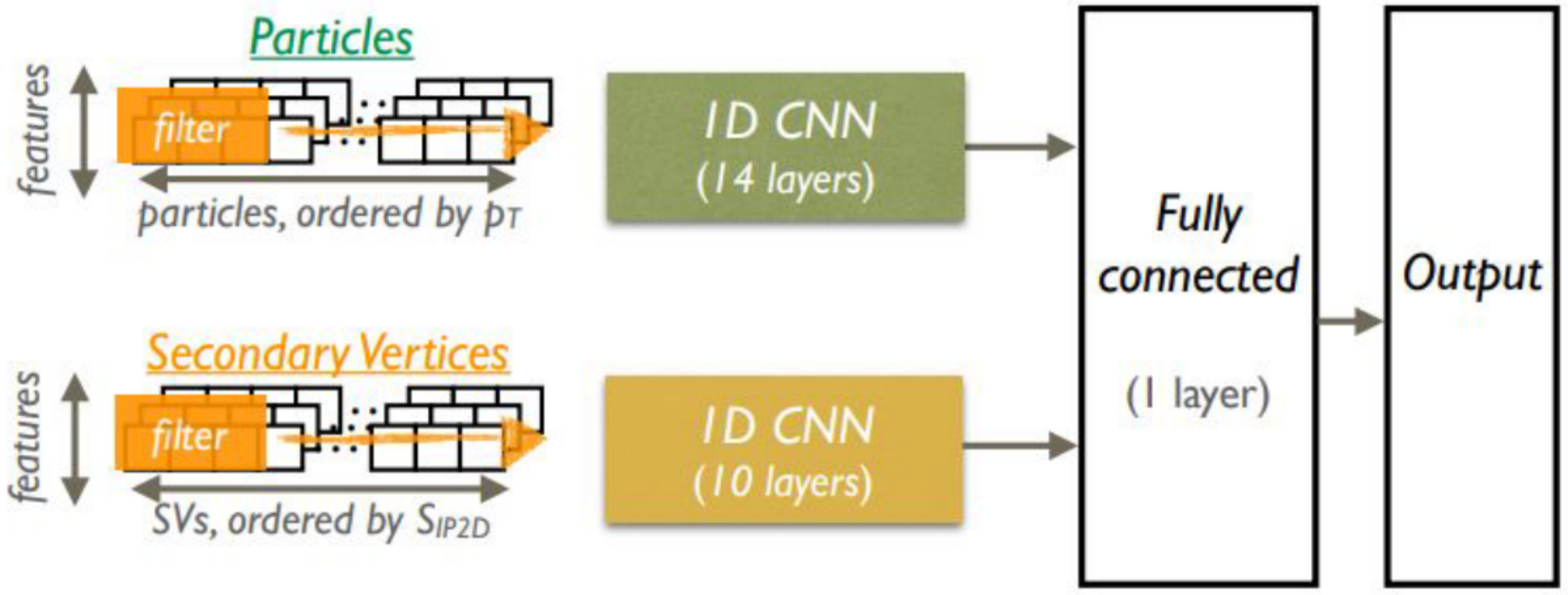



# AK8 jets tagging - DeepAK8

- The DeepAK8 algorithm is a multiclassifier to discriminate jets from W/Z/H/t/other, the main classes are subdivided to minor categories, e.g. Z->bb, Z->cc.
- Model with CNN layers

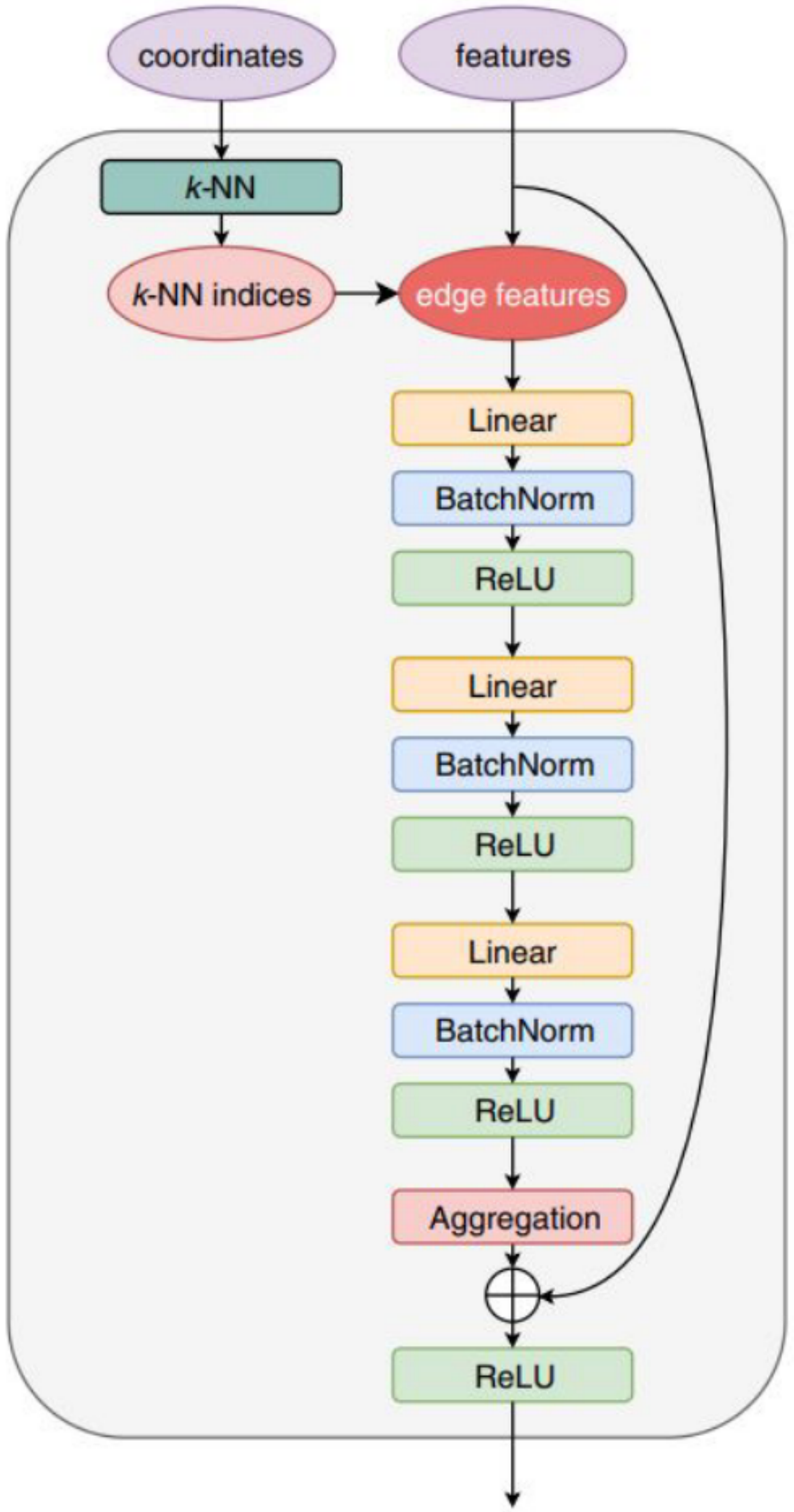
- Two list of input for each jet:

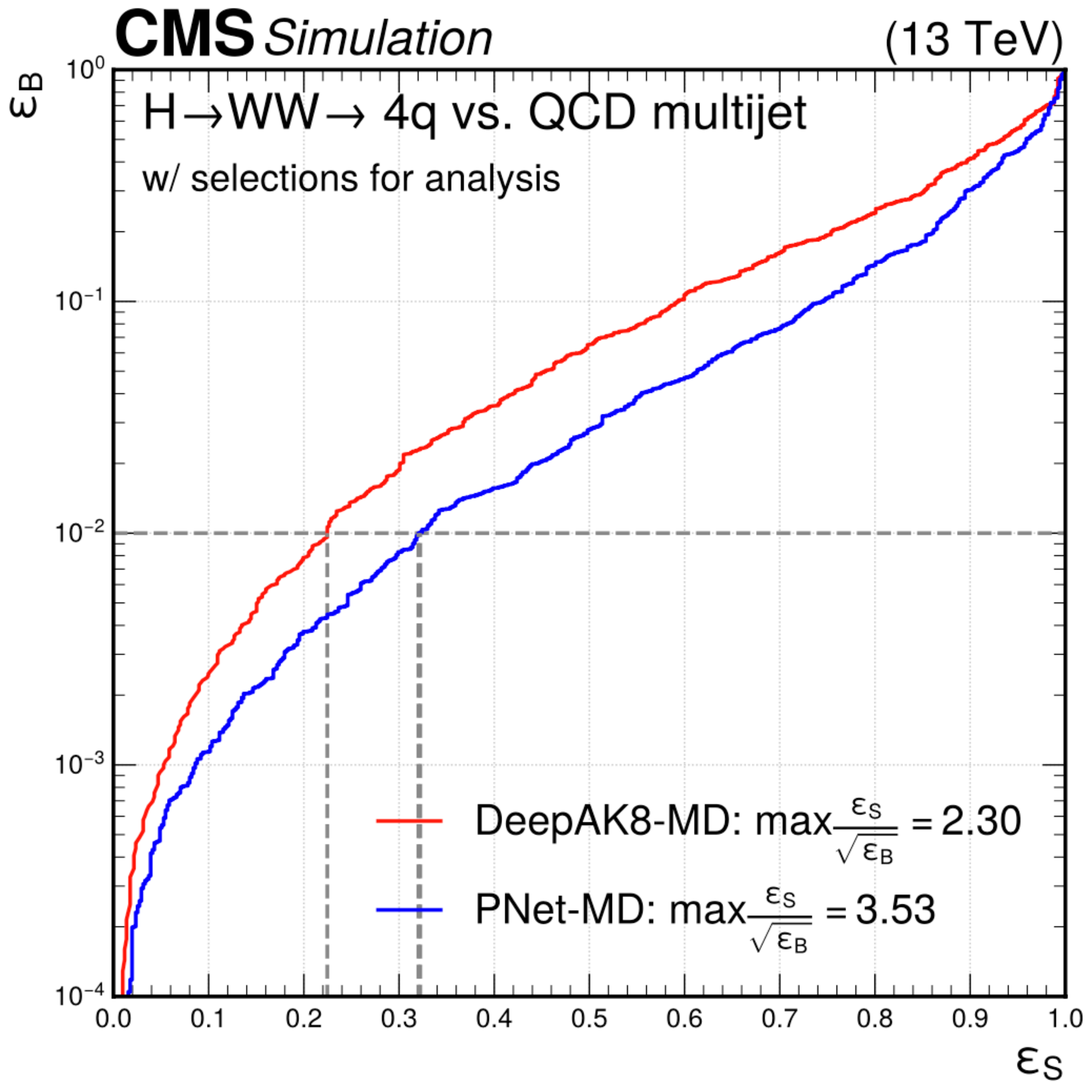
1. 100 jet constituents, order by decreasing transverse momentum, 15 features for each particle
2. Up to 7 SecondaryVertexes (SVs), ordered by 2D secondary ImpactParameter significance, 15 features for each SV



# AK8 jets tagging – ParticleNet

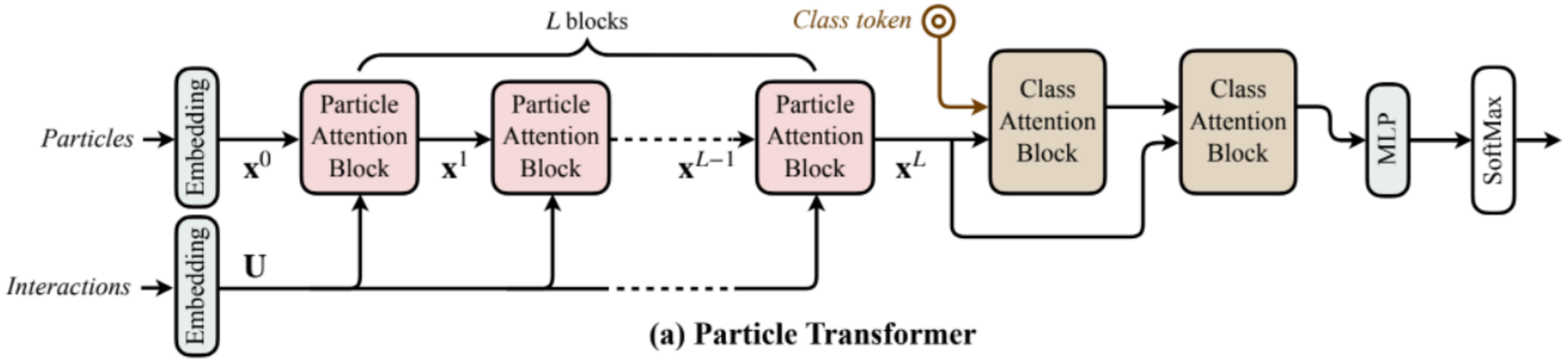
- The ParticleNet algorithm is a Dynamic Graph Convolutional Neural Network
- The jet can be seen as a ParticleCloud, unordered set of entities irregularly distributed. The elements are correlated and with a rich internal structure, taking into account non additive features.
- Imaged based to particle based:
  - no loss of information from pixelation
  - straightforward to include any kind of features for each particle
  - constituent particles of a jet are intrinsically unordered — permutation symmetry!
- Better than DeepAK8!



# AK8 jets tagging – ParticleTransformer

- The ParticleNet algorithm is an attention based Neural Network
- Input embedding: Not only inject single particle information, but also include pair-wise feature
- Better than ParticleNet!
- Updating:
  - Implementation in the CMSSW
  - GloParT3 -> transfer learning with all classes



(a) Particle Transformer


| | All classes Accuracy | All classes AUC | $H \to b\bar{b}$ Rej$_{50\%}$ | $H \to c\bar{c}$ Rej$_{50\%}$ | $H \to gg$ Rej$_{50\%}$ | $H \to 4q$ Rej$_{50\%}$ | $H \to \ell\nu qq'$ Rej$_{99\%}$ | $t \to bqq'$ Rej$_{50\%}$ | $t \to b\ell\nu$ Rej$_{99.5\%}$ | $W \to qq'$ Rej$_{50\%}$ | $Z \to q\bar{q}$ Rej$_{50\%}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PFN | 0.772 | 0.9714 | 2924 | 841 | 75 | 198 | 265 | 797 | 721 | 189 | 159 |
| P-CNN | 0.809 | 0.9789 | 4890 | 1276 | 88 | 474 | 947 | 2907 | 2304 | 241 | 204 |
| ParticleNet | 0.844 | 0.9849 | 7634 | 2475 | 104 | 954 | 3339 | 10526 | 11173 | 347 | 283 |
| **ParT** | **0.861** | **0.9877** | **10638** | **4149** | **123** | **1864** | **5479** | **32787** | **15873** | **543** | **402** |
| ParT (plain) | 0.849 | 0.9859 | 9569 | 2911 | 112 | 1185 | 3868 | 17699 | 12987 | 384 | 311 |

# Calibration tagger

## Model Performance discrepancy from data and MC

- Need to calibrate MC to data to for consistent performance
- SFsBDT method:
  - Use in boosted jet flavor tagging like X->bb, X->cc
  - Use in proxy jets from gluon splitting to a pair of bottom or charm quarks.
  - Can't use for multiple subjets cases (>=3)
- Lund jet plane reweighting method:
  - Use primary lund-plane regions to extract reweighting factors for each subjet
  - Works for multiple subjets cases
  - large systematic uncertainties -> no harm for search analysis