

DeepFlavour: Deep Neural Networks for Flavour Tagging

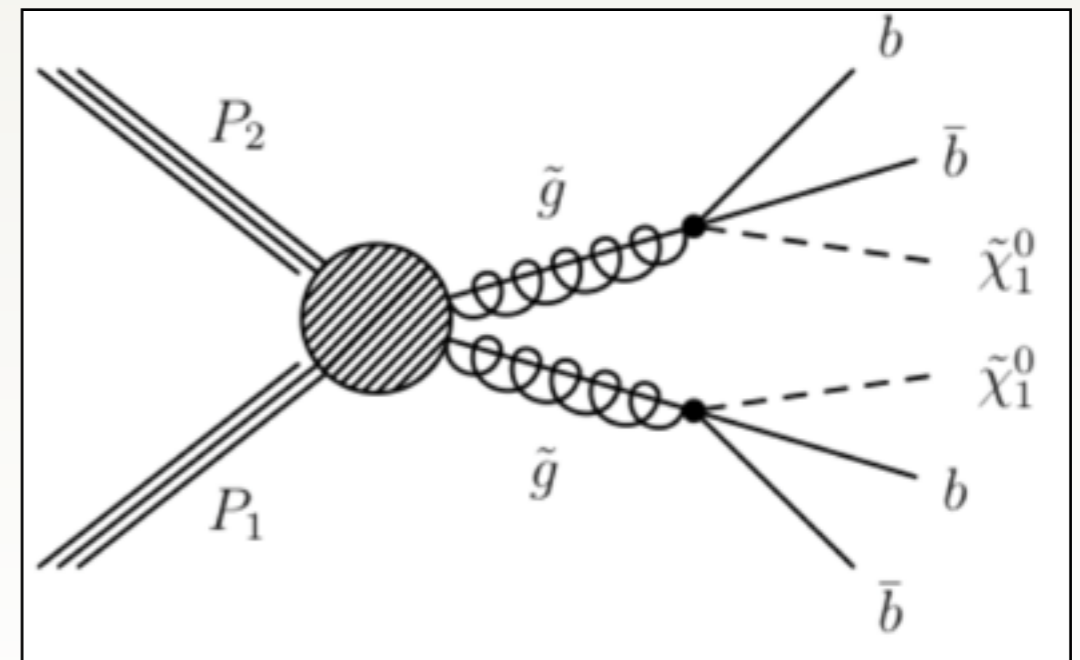
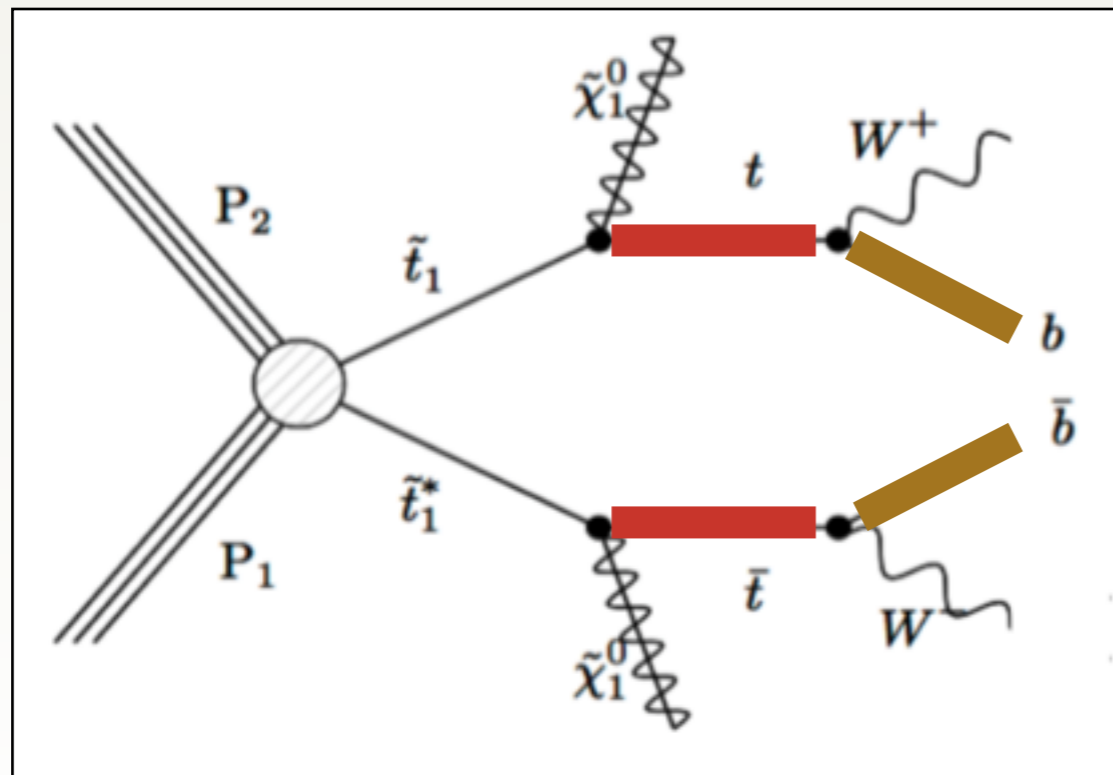
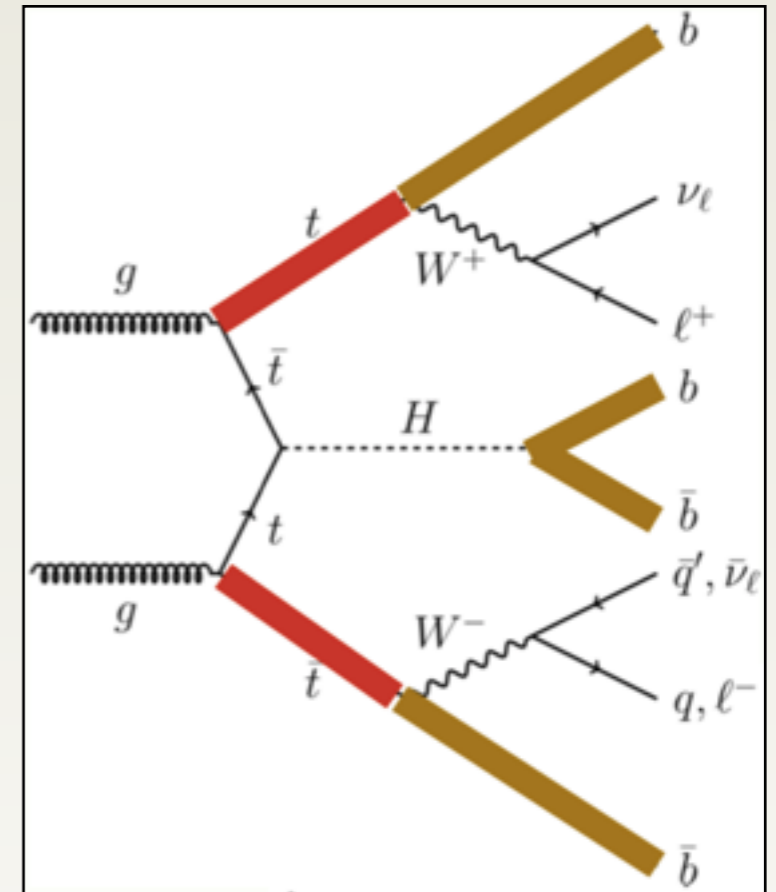
Jan Kieseler

10.5.2017



(Heavy) Flavour Identification

- Important for various SM and BSM analyses w.r.t top quarks
($\text{BR}(t \rightarrow Wb) \approx 100\%$)
 - Precision top-quark measurements
 - BSM physics with strong couplings to top quarks
 - Veto against SM top quarks
- But also important for other processes:
 Hbb , $V+bb$, ..., $V+cc$
- All form (b)jets in the final state

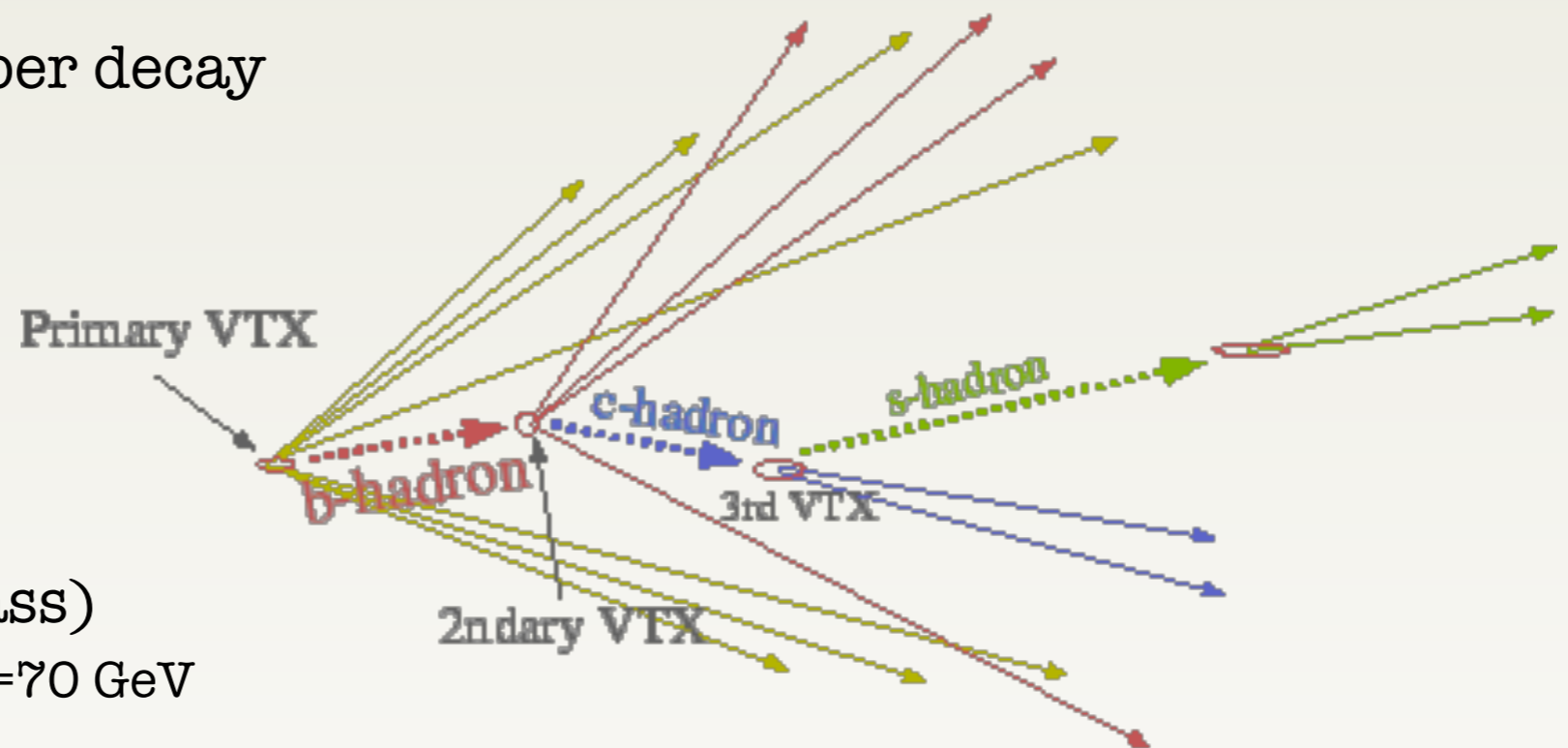


Long-lived B hadrons

- Large mass ~ 5 GeV
- Long lifetime ($c\tau \approx 450\mu\text{m}$)
- Around 5 charged particles per decay
- Leptons from EWK decays
- Tertiary vertex ($b \rightarrow c$)

In the detector

- Displaced tracks
- Secondary vertices (high mass)
 - about 5mm decay length @ $E(b) = 70$ GeV
- Slightly different jet shapes w.r.t gluon/light jets
- Leptons within the jets



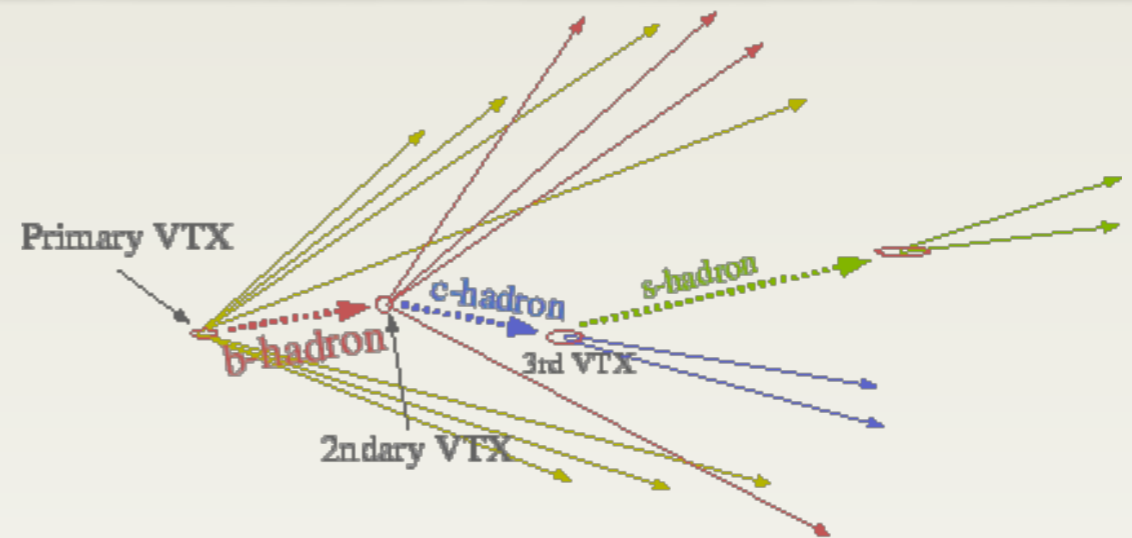
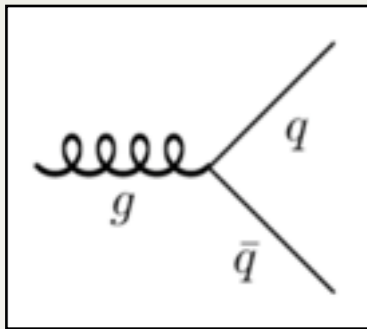
[<http://acfahep.kek.jp>]

➡ Features used in almost all b-identification (tagging) algorithms

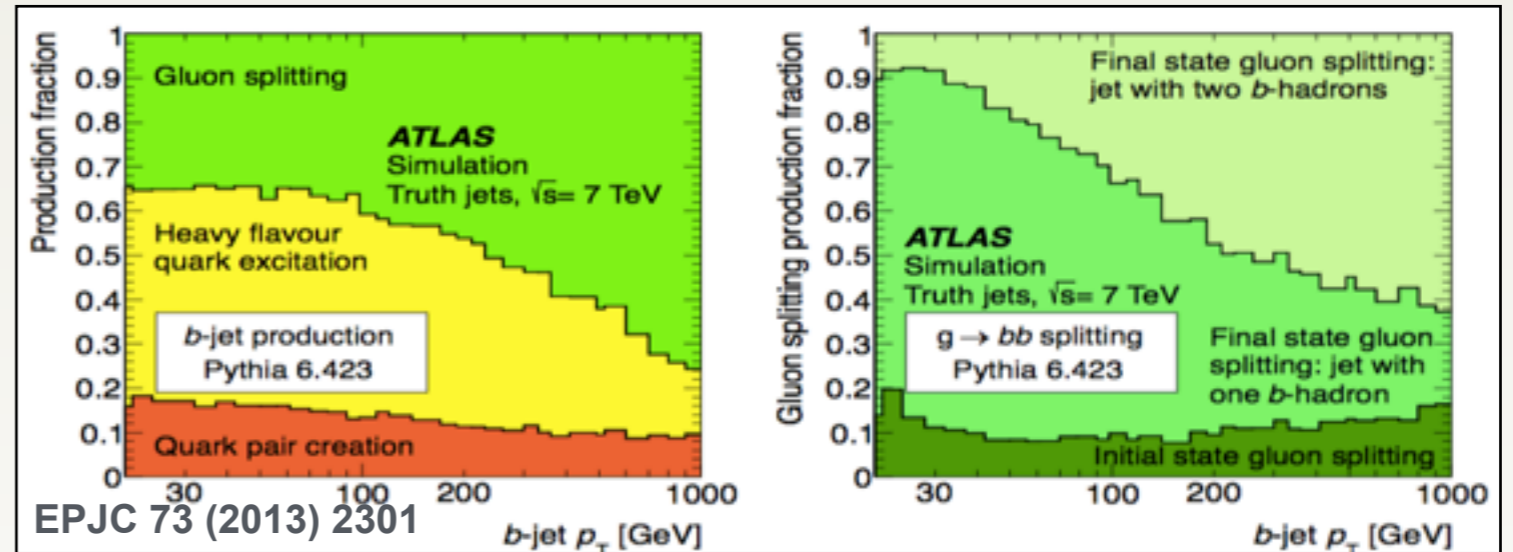
Jet Classification

Physics scenarios:

- Single b-jet with one b-hadron
- Gluon splitting
(dominant fraction in QCD events)



- Boosted bosons to bb
- ‘Real’ c-quark jets



4 mutually exclusive categories (multi-classification problem):

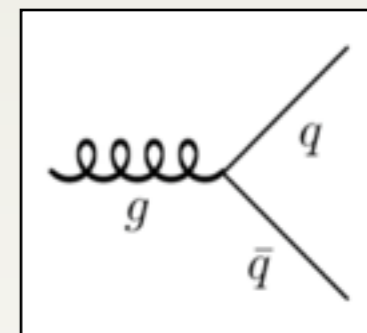
- Exactly 1 b-hadron
- Exactly one c-hadron (but no b-hadron)
- Two or more b-hadrons
- Light jets (udsg) (no q/g discrimination for now)

DNN

- Good classification performance
- Fast application
- Highly scalable
- Needs a generic training

Training samples (40 - 50 M jets):

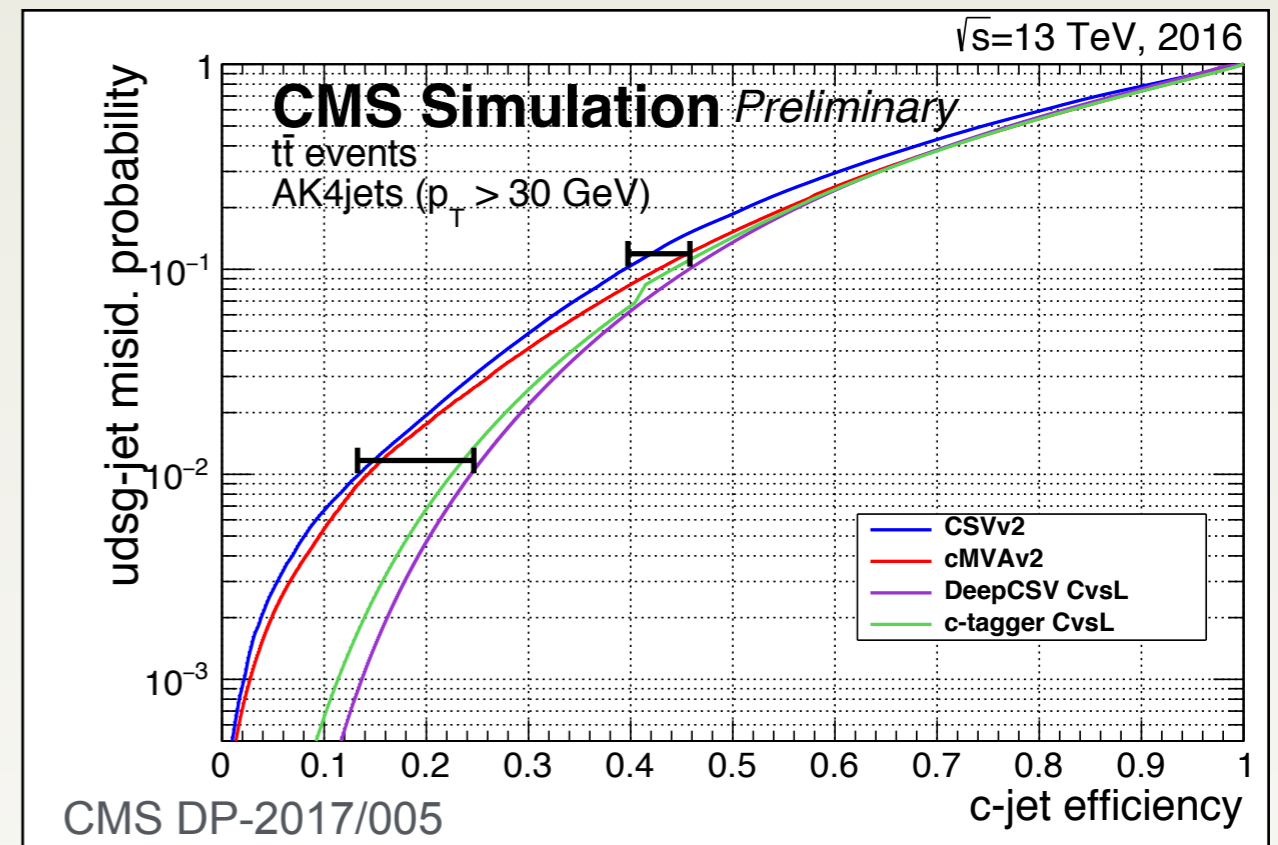
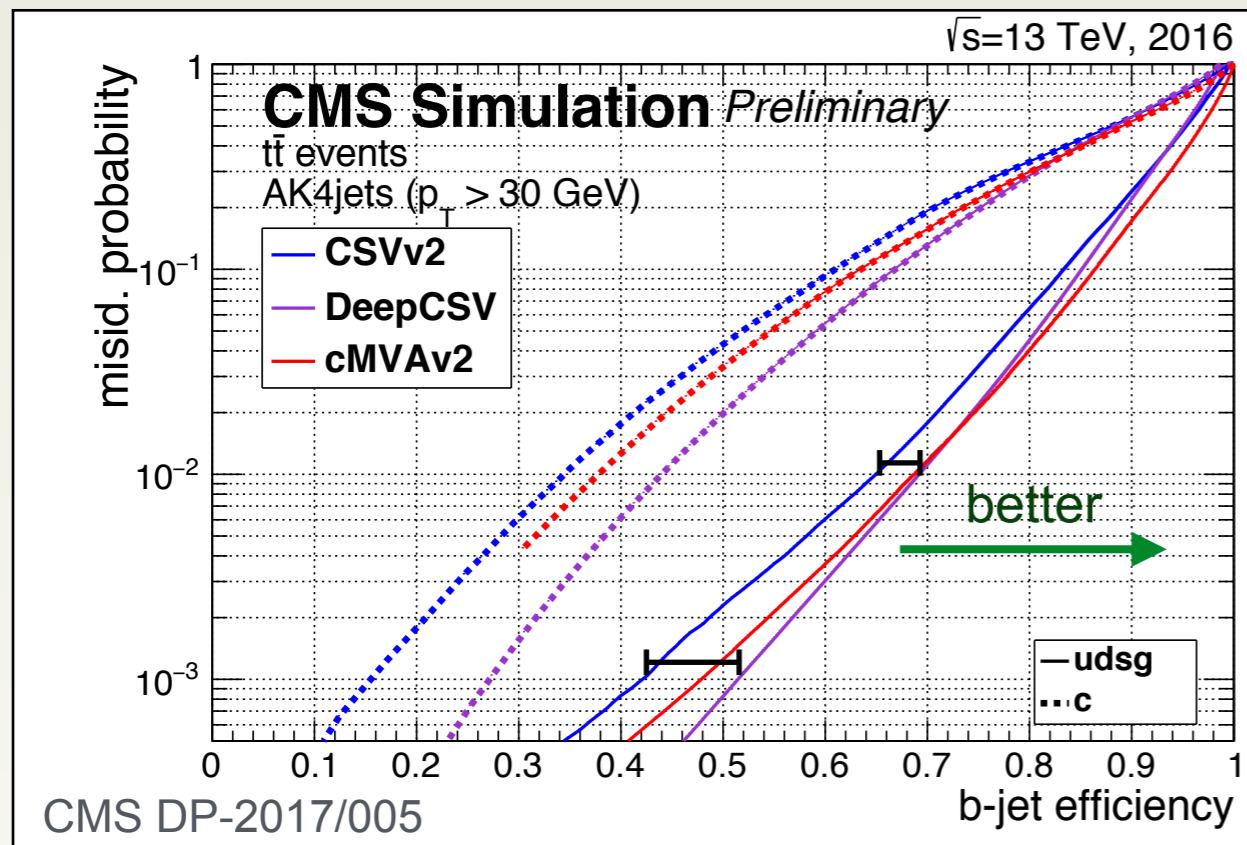
- QCD:
 - ▶ very clean, almost no overlaps with (prompt) leptons
 - ▶ good source of gluon splitting sample, flavour excitation, flavour creation
- Top-quark pair events:
 - ▶ includes accidental overlap of leptons
 - ▶ b quarks from top decays and c quarks from W bosons



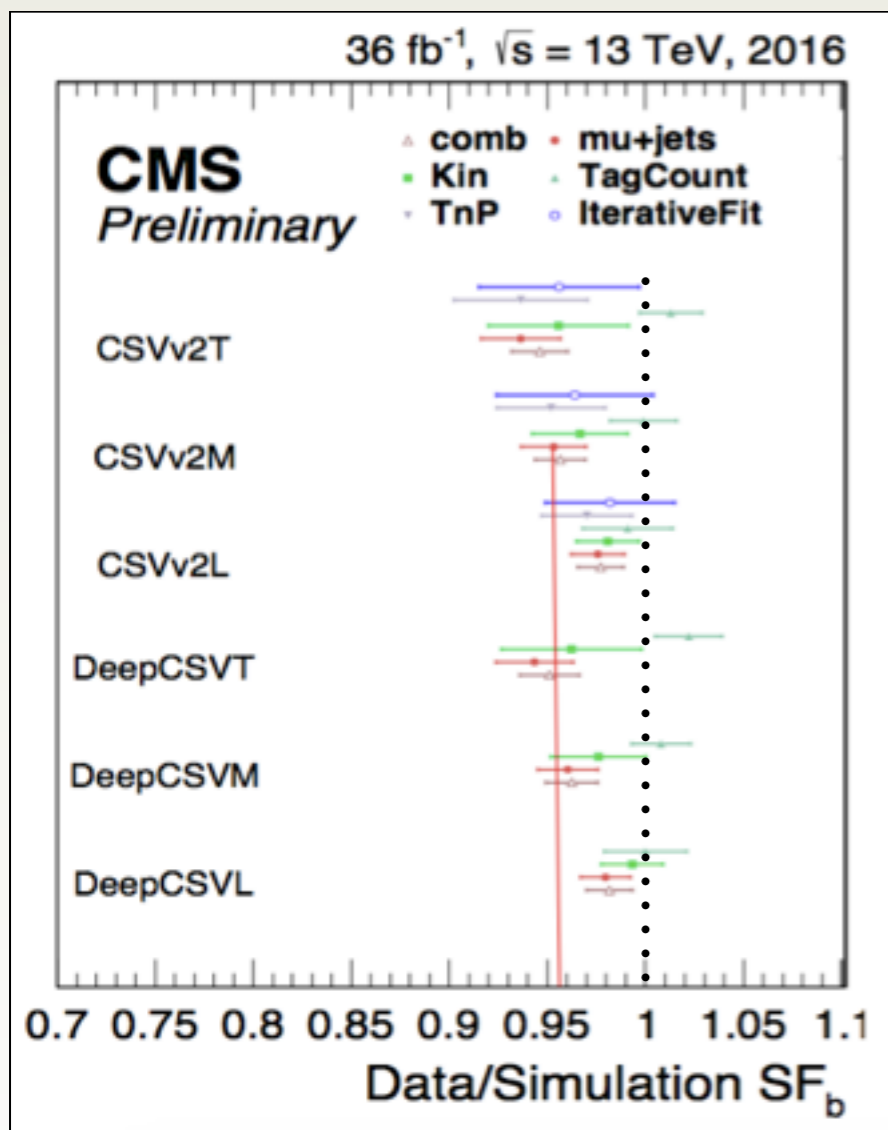
DeepCSV

- Fully commissioned DNN based tagger in CMS used for analyses in 2016

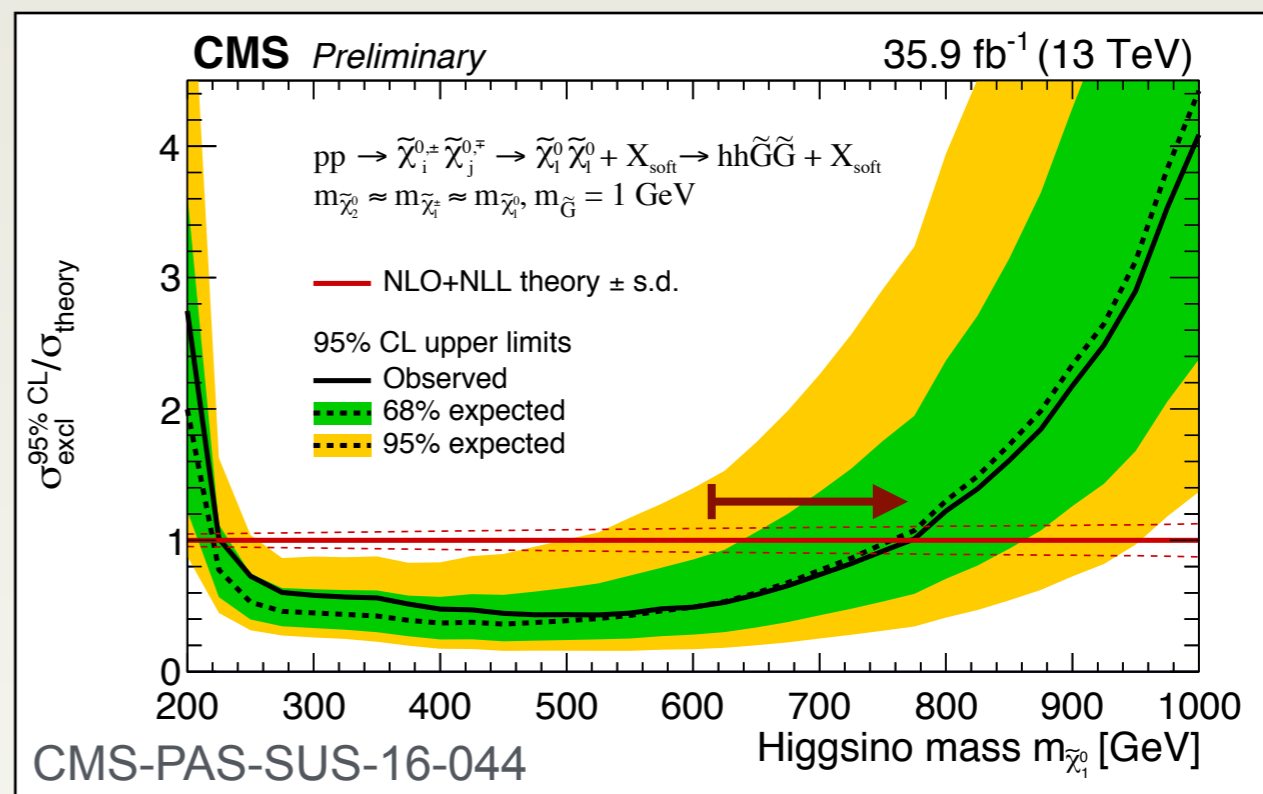
DeepCSV Performance (2016)



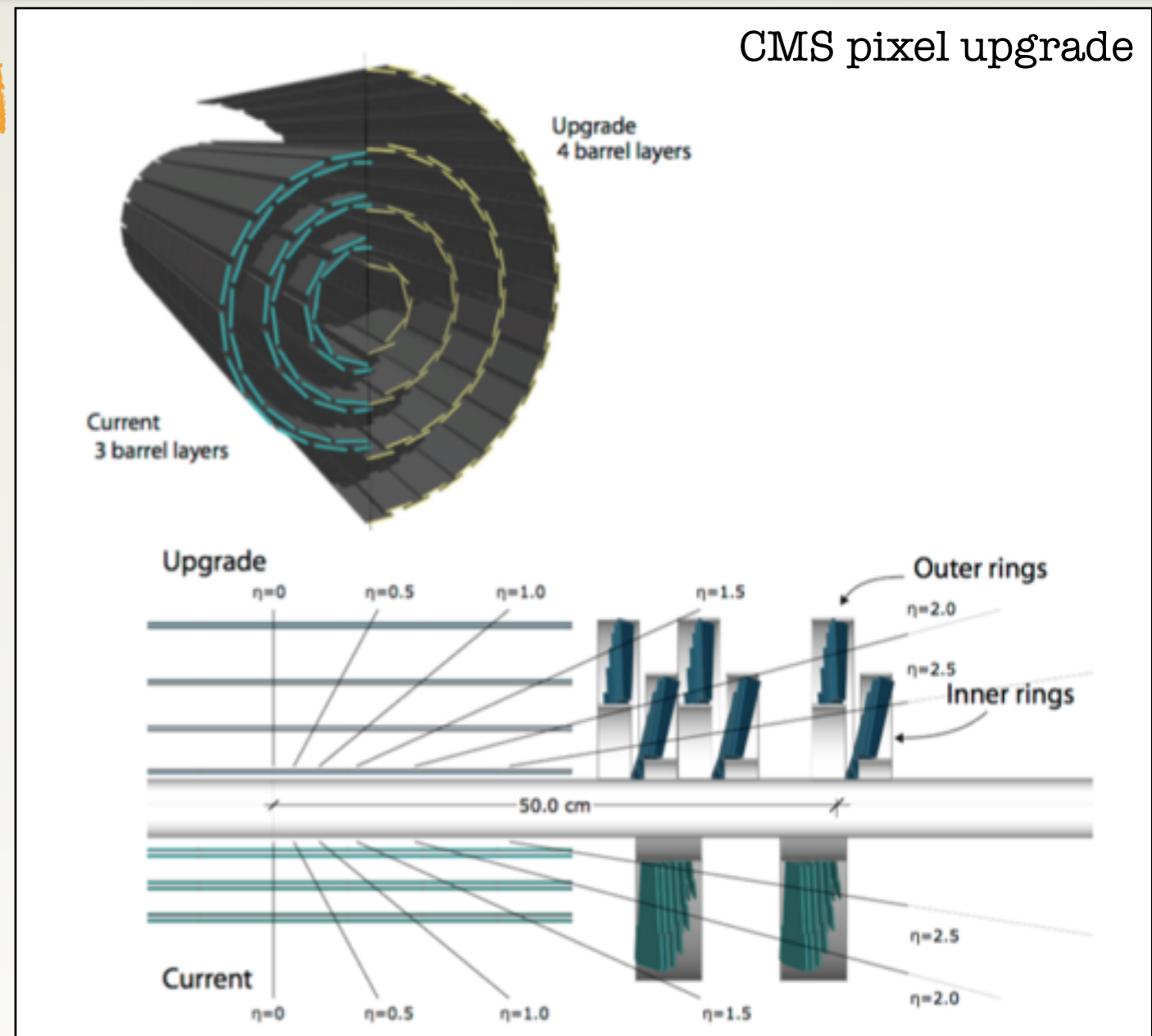
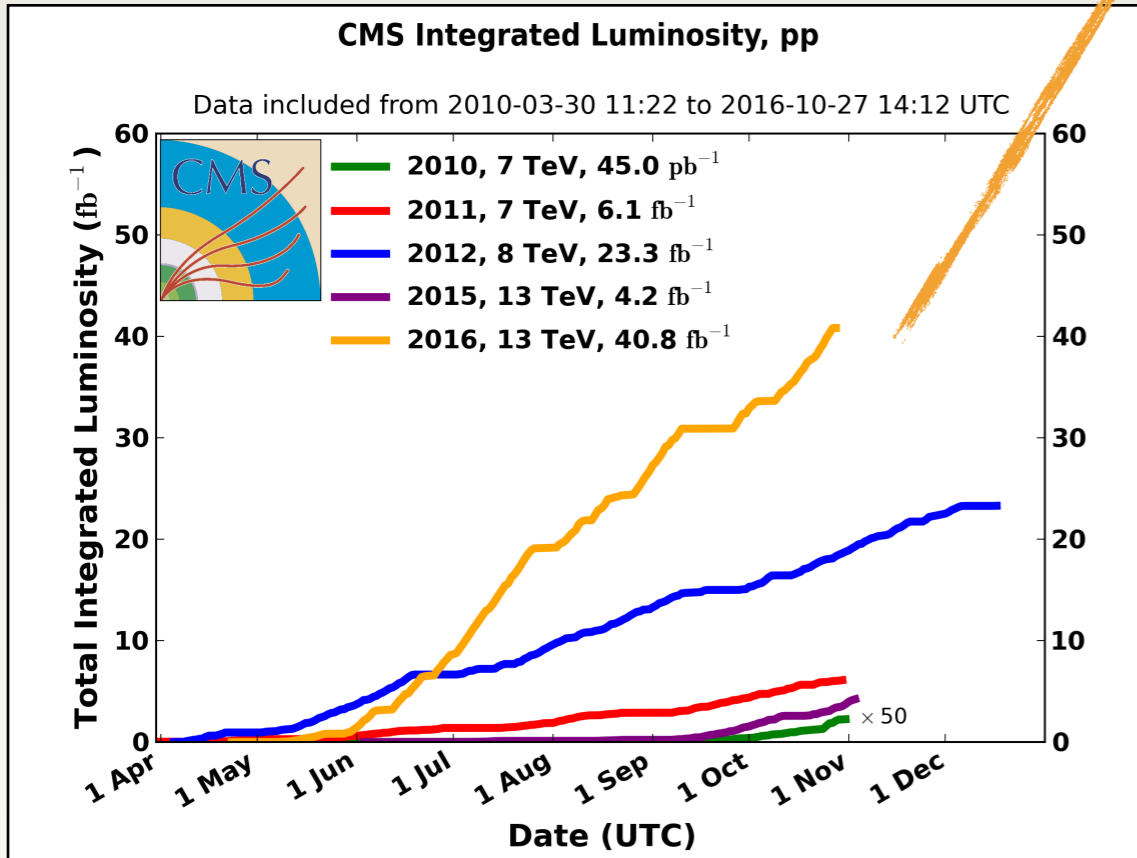
- Uses same input as CSVv2 (slightly extended)
- DeepCSV outperforms other taggers
 - In particular for high-purity selection (low misid. probability)
 - Also performs very well in terms of c-tagging



CMS DP-2017/005

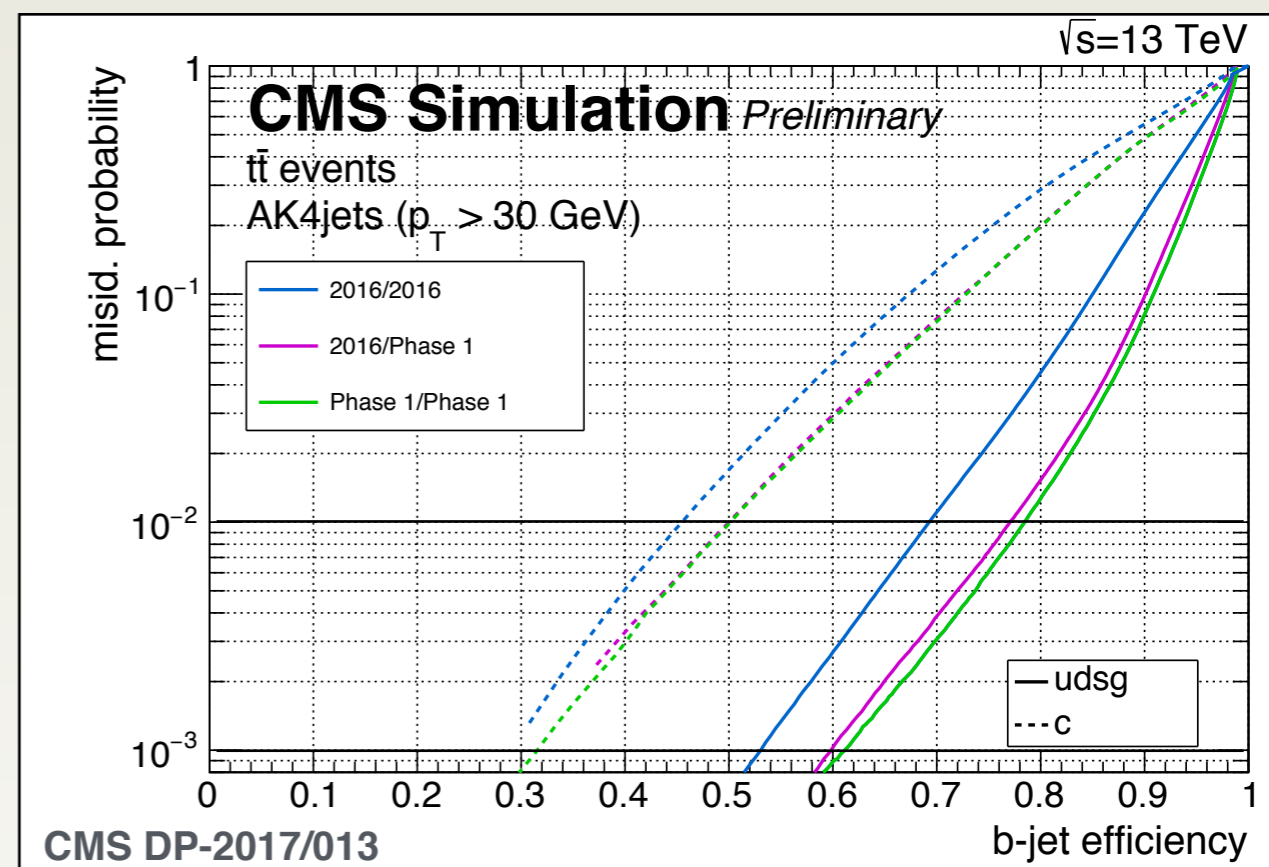
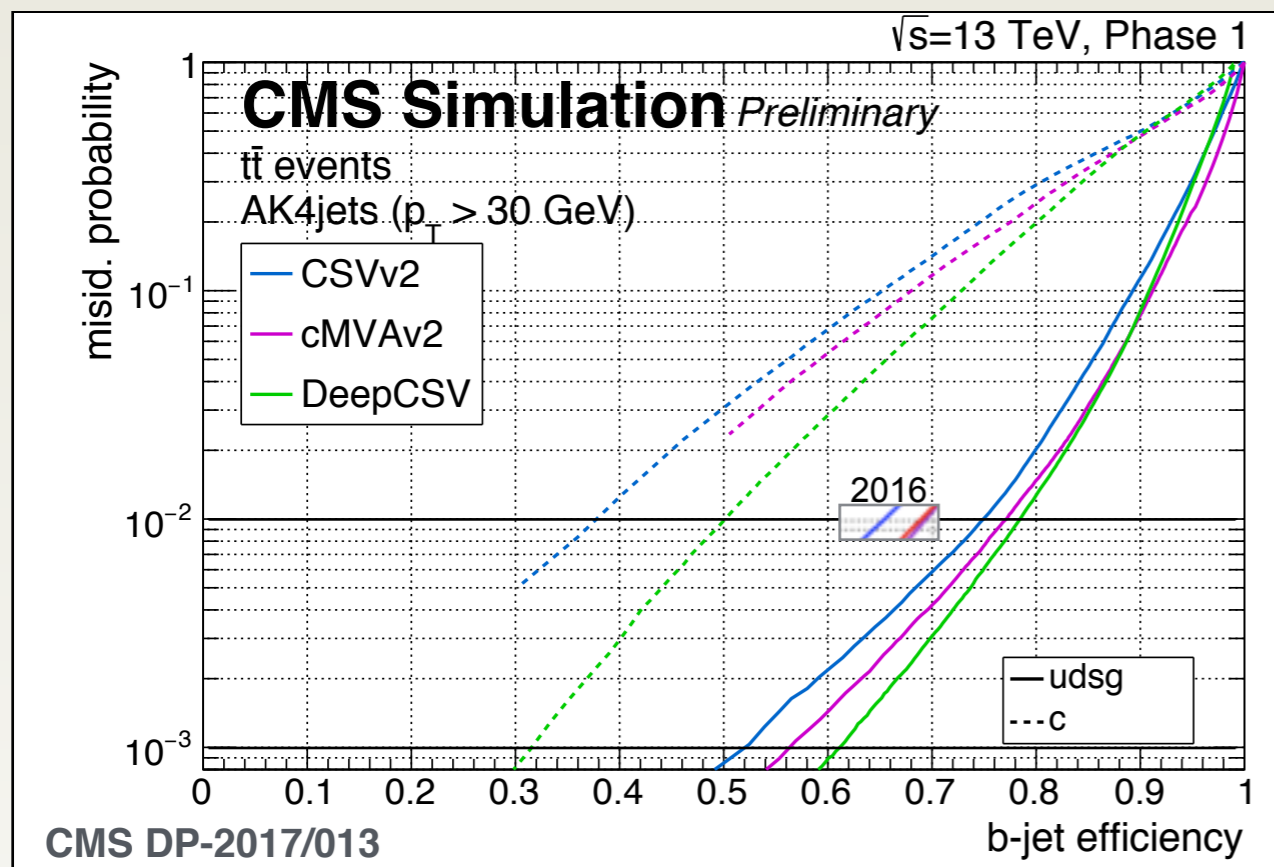


- Fully commissioned on 2016 data
- Simulation well described, only small corrections
- Used in physics analyses
 - ▶ e.g. 2,3,4 b-jets categorised Higgsino search
 - ▶ Significant gain



- Now installed, data taking will start very soon
- Gain in resolution for secondary vertices, displaced tracks etc.
- ➔ Strong impact on b-tagging expected

DeepCSV with the new Detector Design



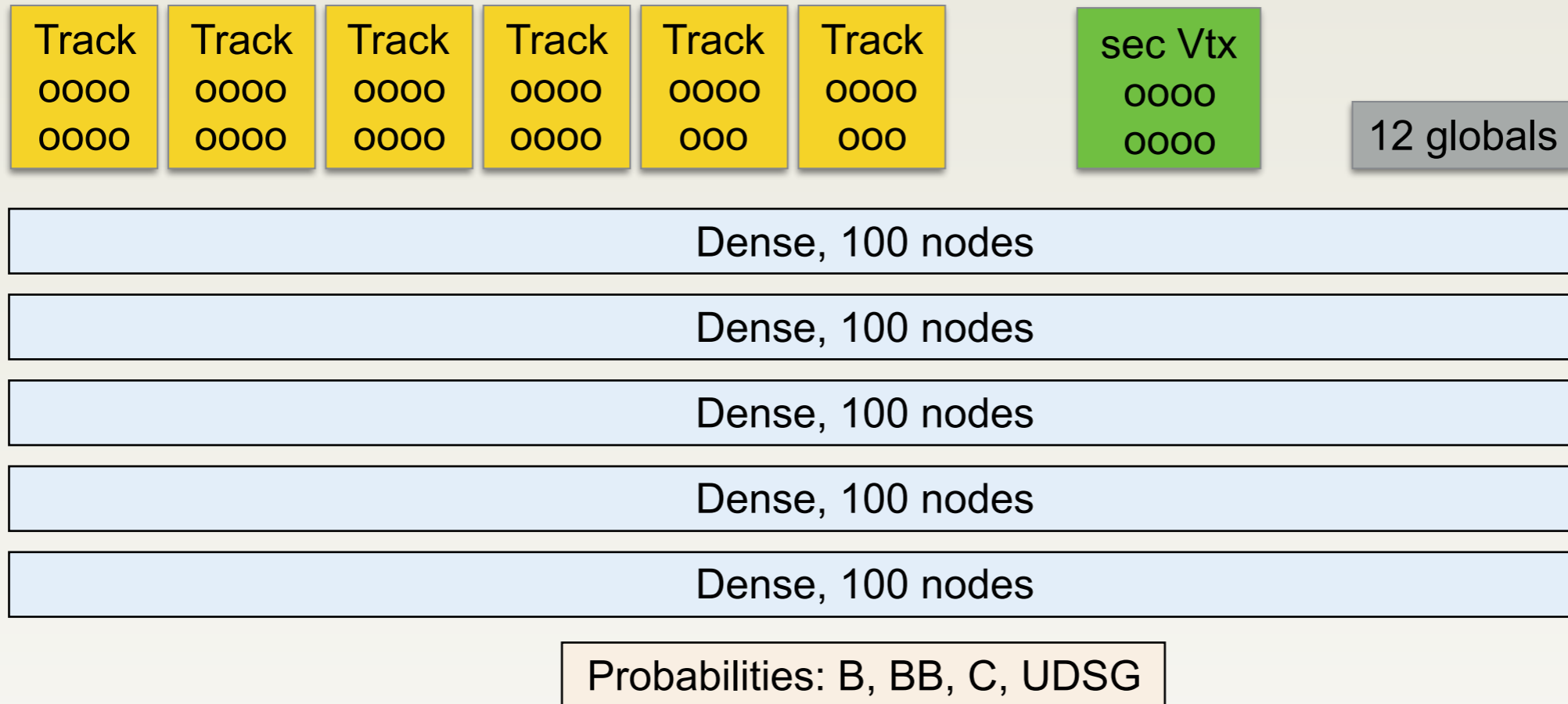
- Clear benefit from the new pixel detector for all taggers

DeepCSV

- Almost 80%* b vs. light flavour discrimination efficiency for 1% misid. prob.
- Re-training of the DNN increased performance even (slightly) more
 - Re-trained on 40M jets from QCD and top-quark pair events
- First successful step towards DNN-based taggers

* small degradation expected with improved simulation

DeepCSV: Details



- Uses ‘standard’ input to also other CMS b-tagging algorithms: **66** inputs
 - ▶ up to 6 strongly preselected charged particle tracks with 7 properties, each
 - ▶ additionally one variable for the first 4 tracks
 - ▶ up to 1 selected secondary vertex (8 properties)
 - ▶ 12 global per-jet variables
- Sorted by impact parameter significance
- Fully connected dense network (ReLU, last softmax, loss: cross entropy)
 - ▶ ‘simple’ network structure

Convolutional Neural Networks



[1]



convolutional
network

- Used a lot for image processing
- Looking at tiles of an image
 - ▶ Translation invariant: same network for every tile (significantly reduced number of neurons to be trained)
 - ▶ Learns to recognise certain features (e.g. edges, specific colors) that are important for the task
 - ▶ Can compress e.g. 250x250 pixels in each tile to a small set of features

[1] <https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721>

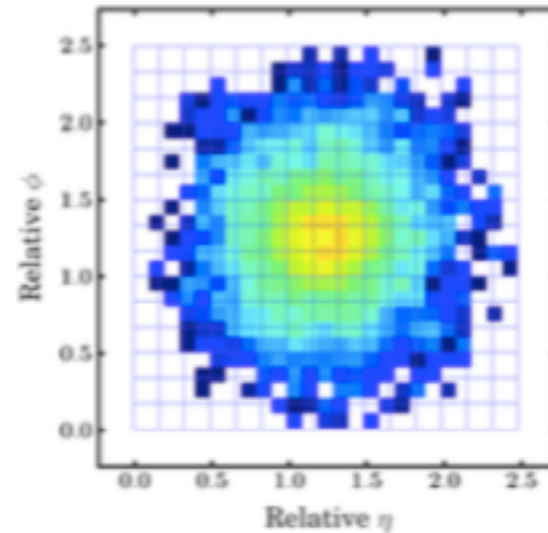
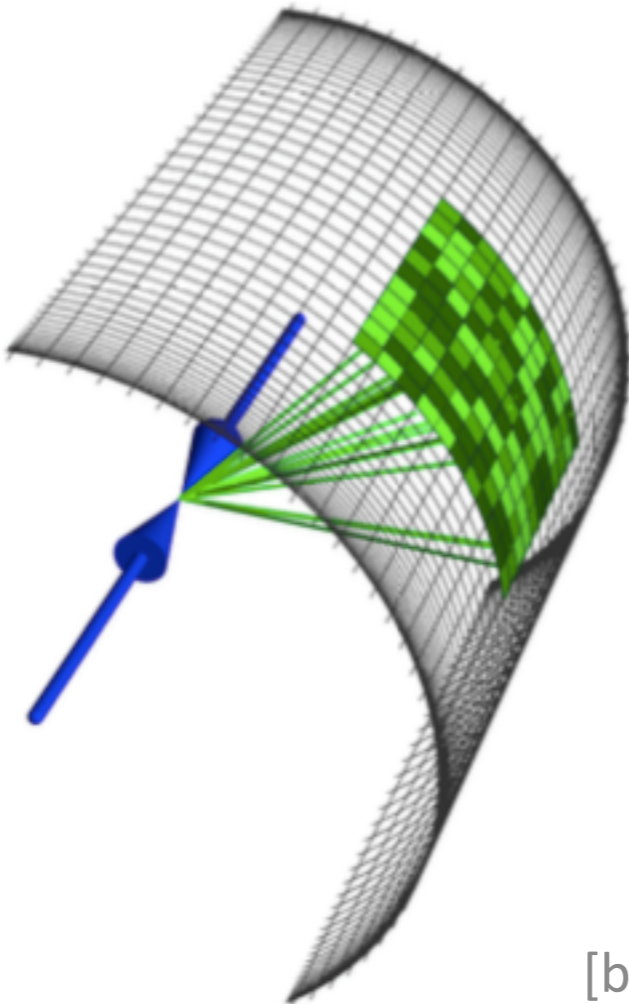
Jet Images

[Almeida et al. arXiv: 1501.05968]

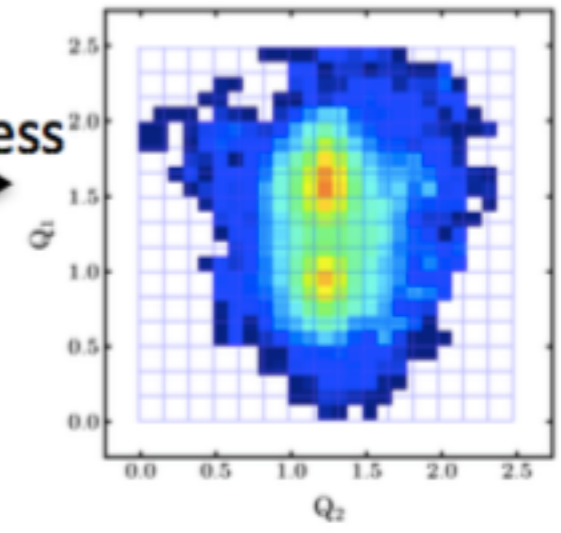
[Cogan et al. arXiv:1407.5675]

- Treat energy deposits as image

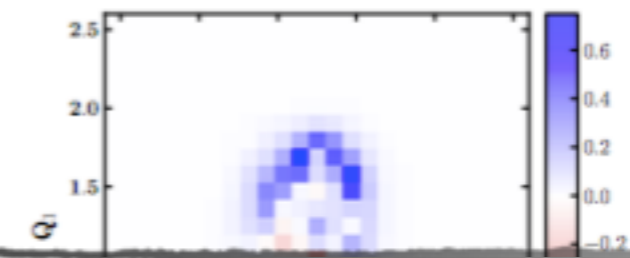
Application to boosted W tagging



preprocess
→

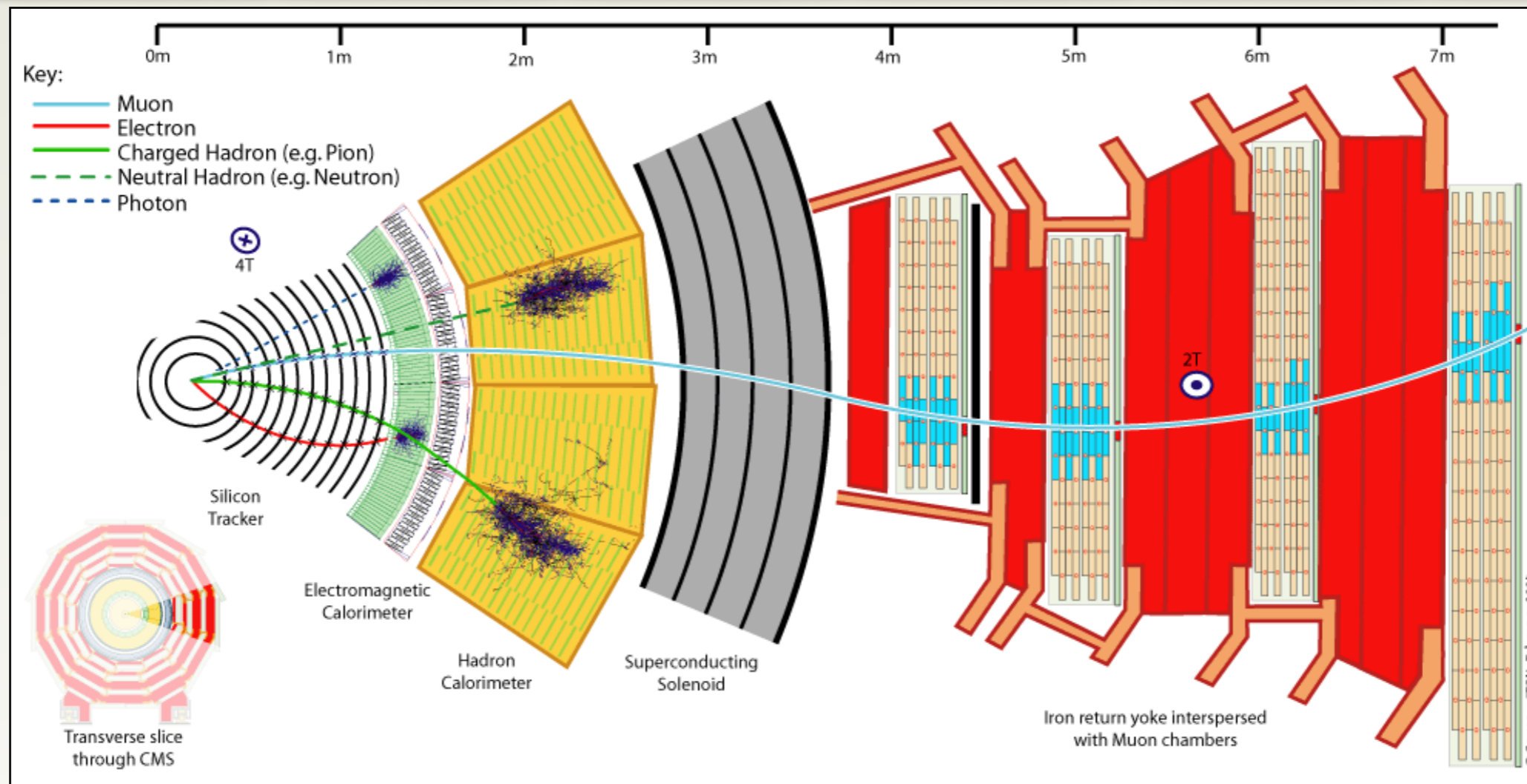


[by Matthew Schwartz]



- Not talking about jet images, here

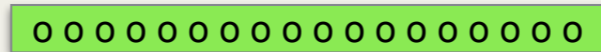
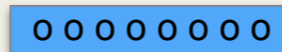
Logical Structures for CNN



- Another equivalent to image tiles in HEP: jet constituents
(In CMS: particle flow candidates)
- Consider 3 categories
 - ▶ Charged particles
 - ▶ Neutral particles
 - ▶ Secondary vertices

CNN input

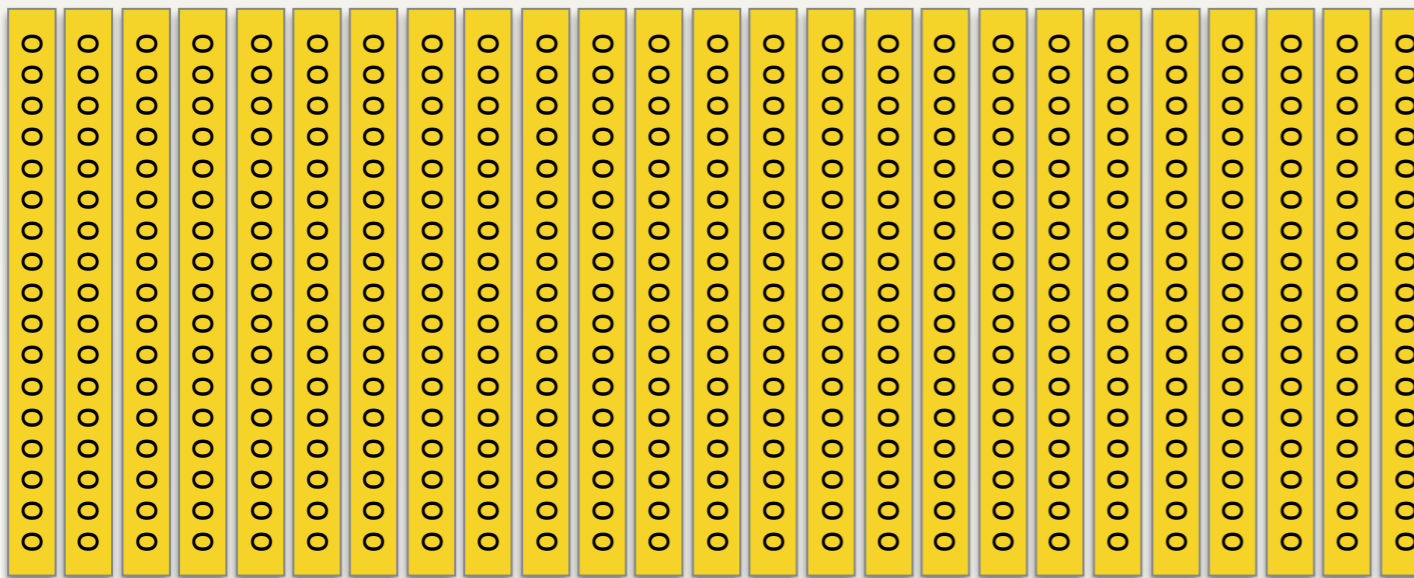
- Define hierarchical 'tile' ordering for each collection
 - Based on measure of displacement
- Select a subset of properties per candidate
 - charged candidate: 16
 - neutral candidate: 8
 - secondary vertex: 17



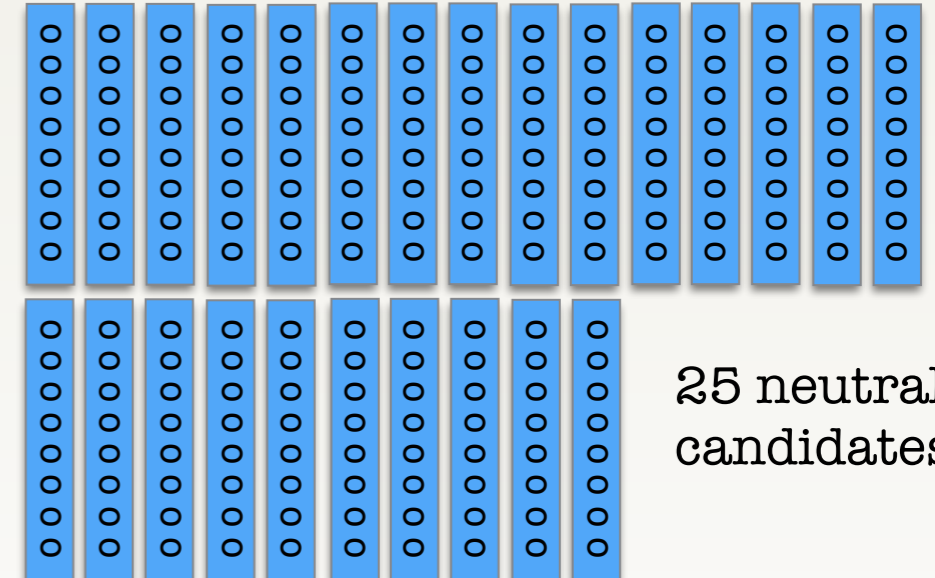
for comparison:
DeepCSV

oooooooooo	12 globals
oooooooooo	oooooooooo
oooooooooo	oooooooooo
oooooooooo	oooooooooo

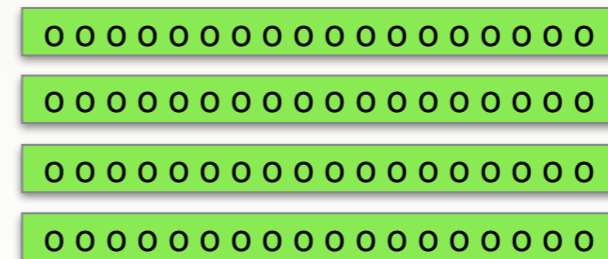
- Final input to the CNNs (~650 inputs in total)



25 charged candidates



25 neutral candidates

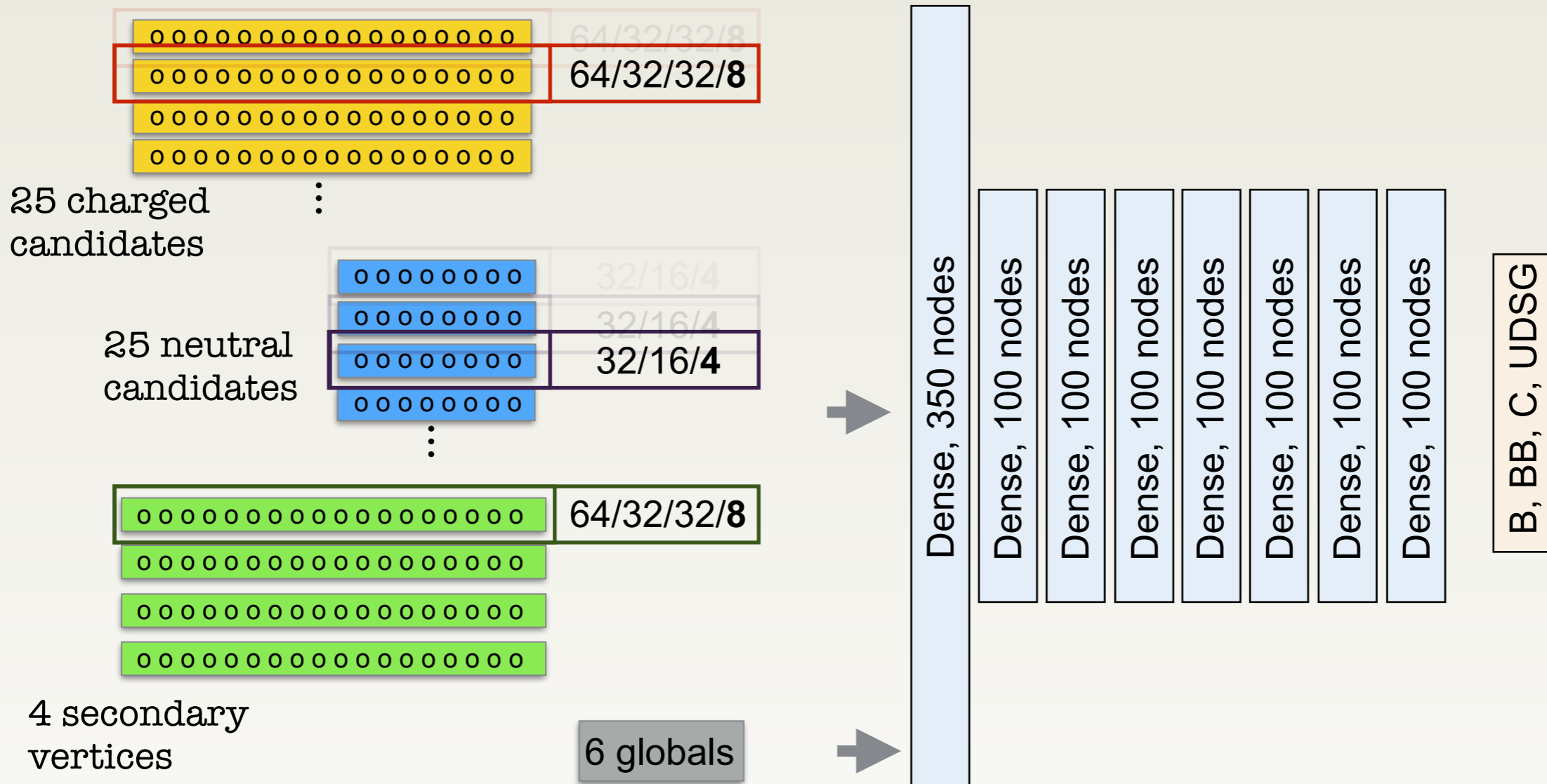


4 secondary vertices

6 globals

➔ Almost full set of jet constituents without strong preselection

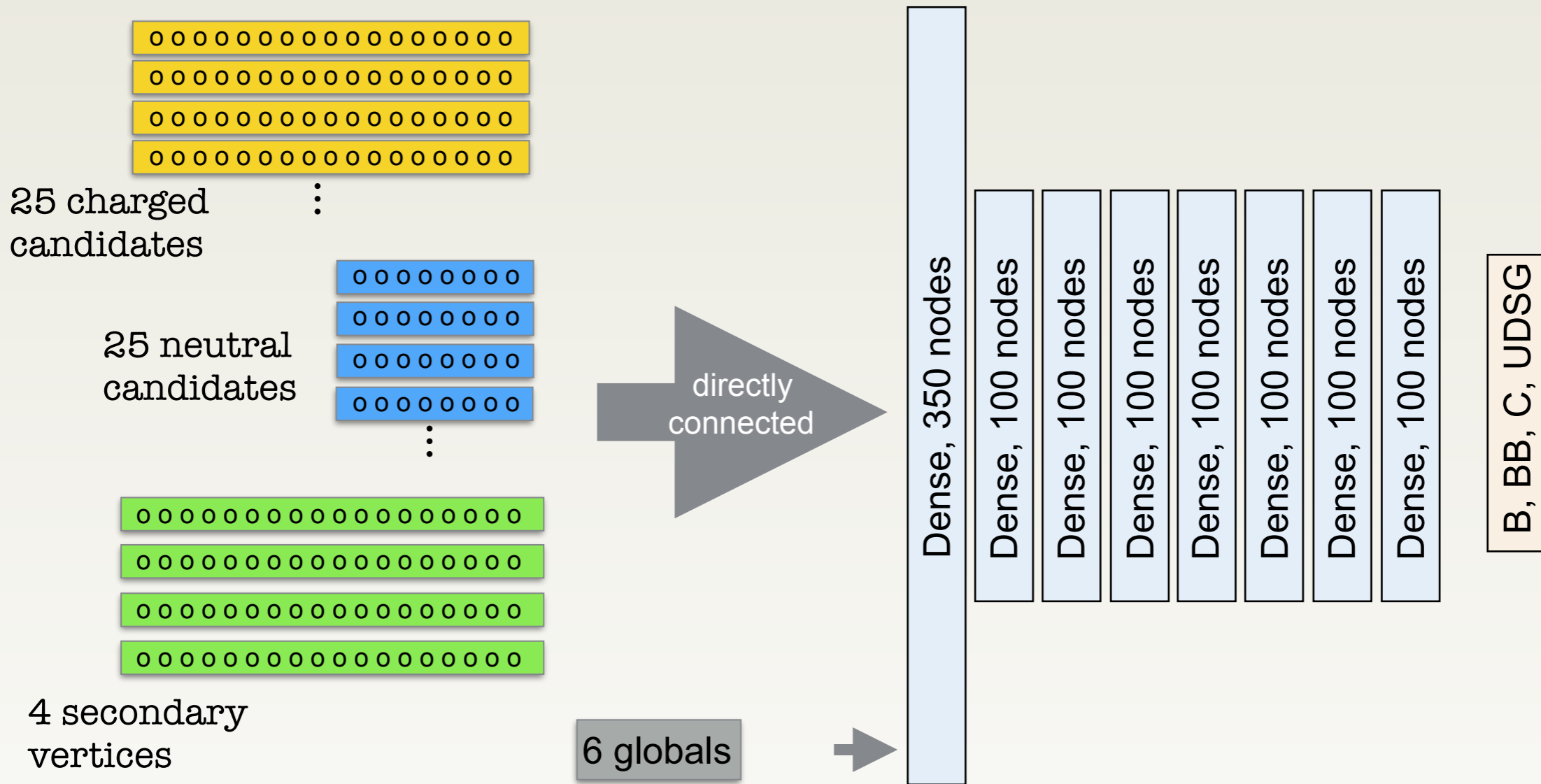
The DeepFlavour DNN*



- Use multiple convolutional layers per collection
 - Feature transformation and compression of about a factor of 2
- Look at each individual particle (1x1 convolution, ReLu)
- Forward CNN output to multiple dense layers (ReLu, last softmax)
- Trained on 40M jets from top-quark pair and QCD events

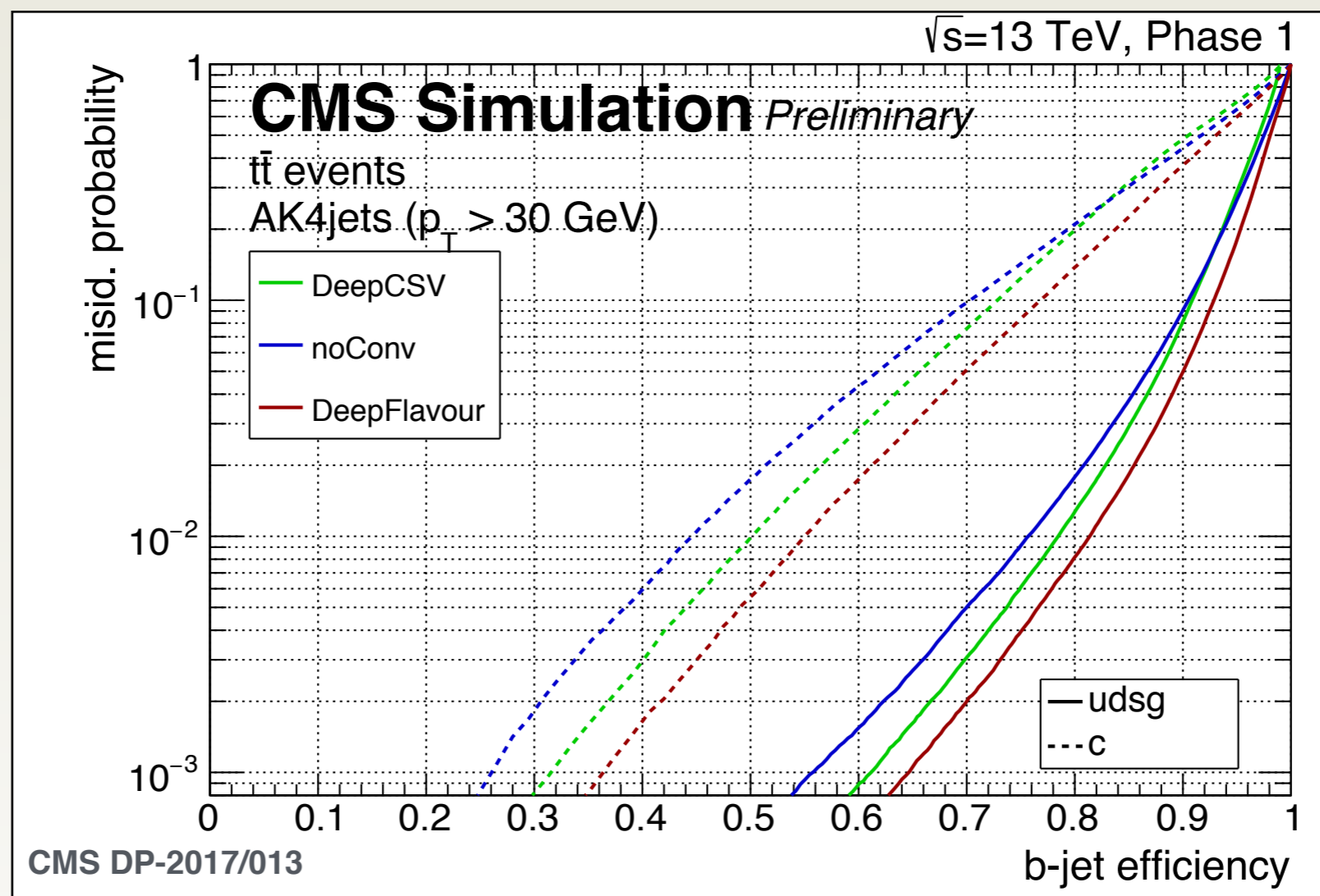
*under development

The noConv DNN



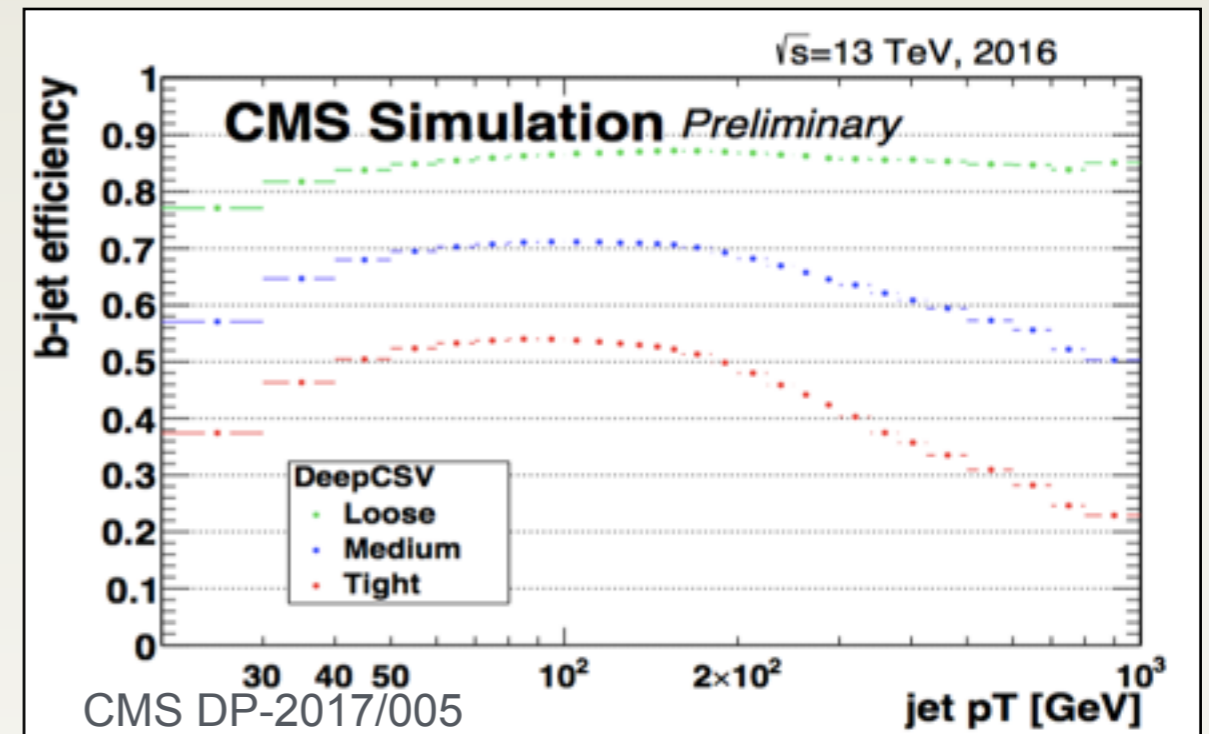
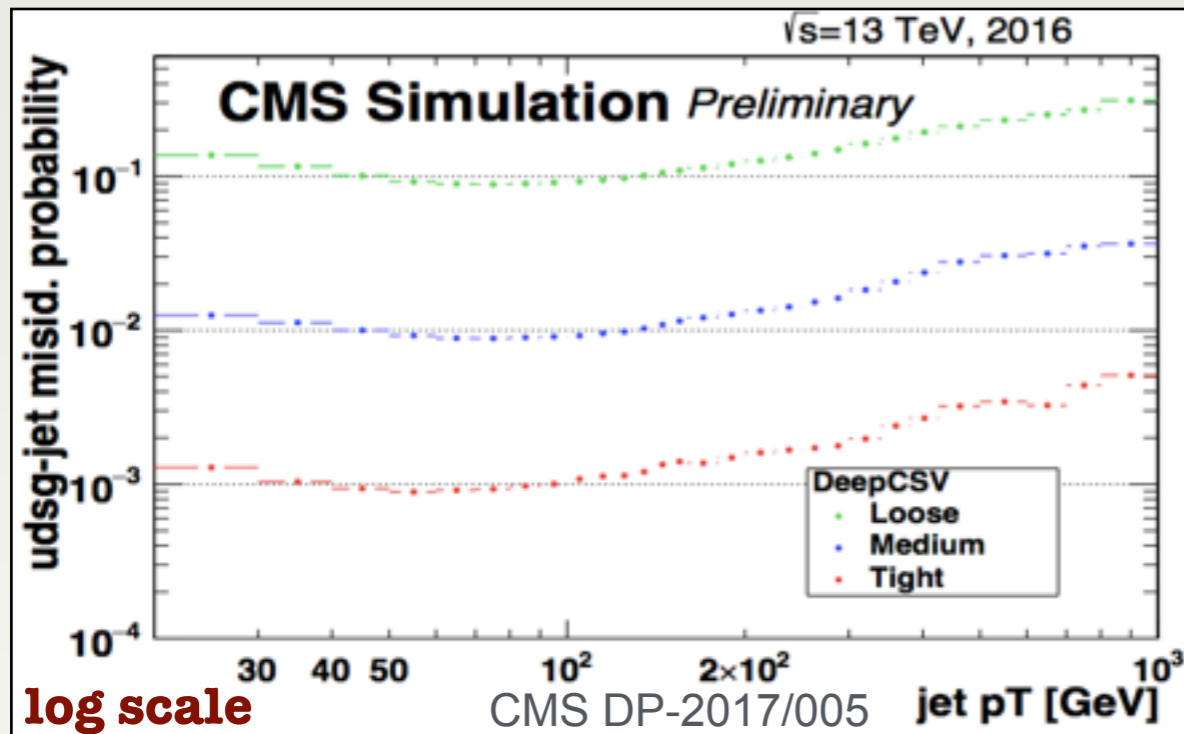
- Removed convolutional layers to estimate effect of property transformation and compression on the network performance
- Only for comparison

DeepFlavour Evolution

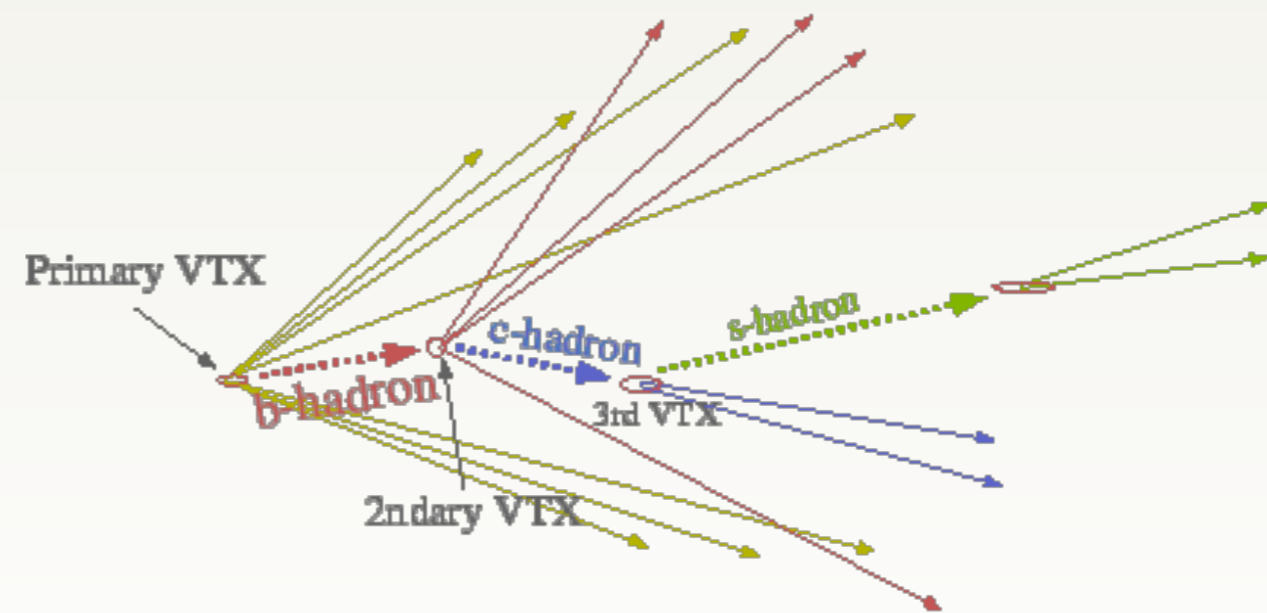


- Simply adding more information degrades performance
- Adding convolutional layers increases the performance significantly
 - Profit from transformation and compression of particle properties

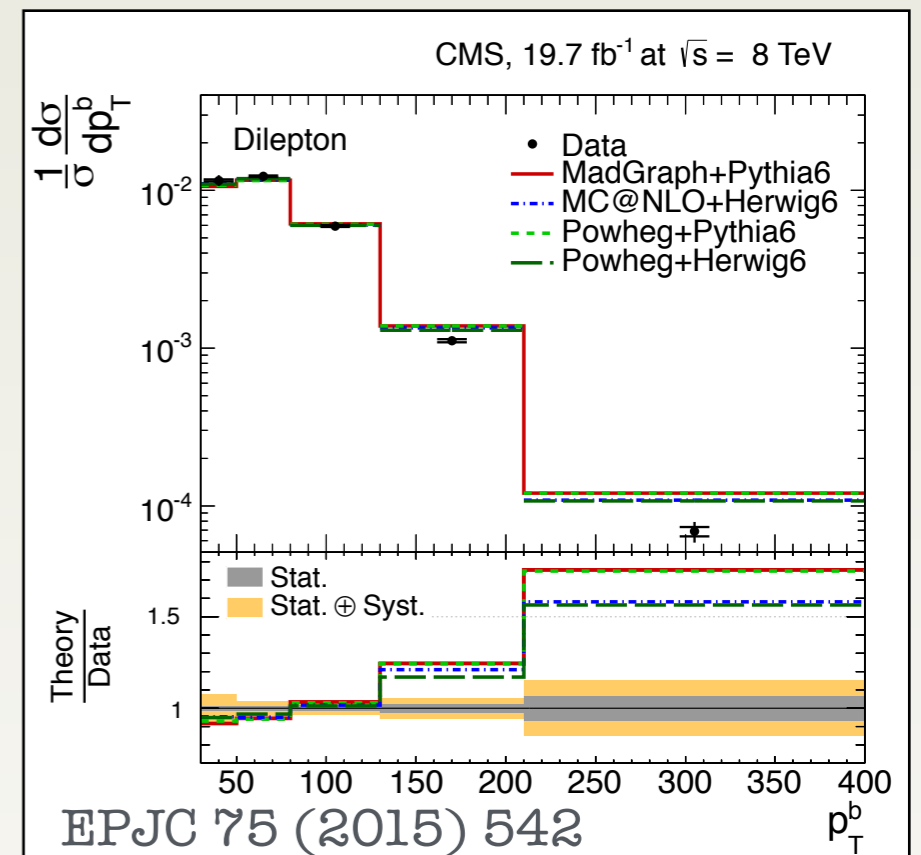
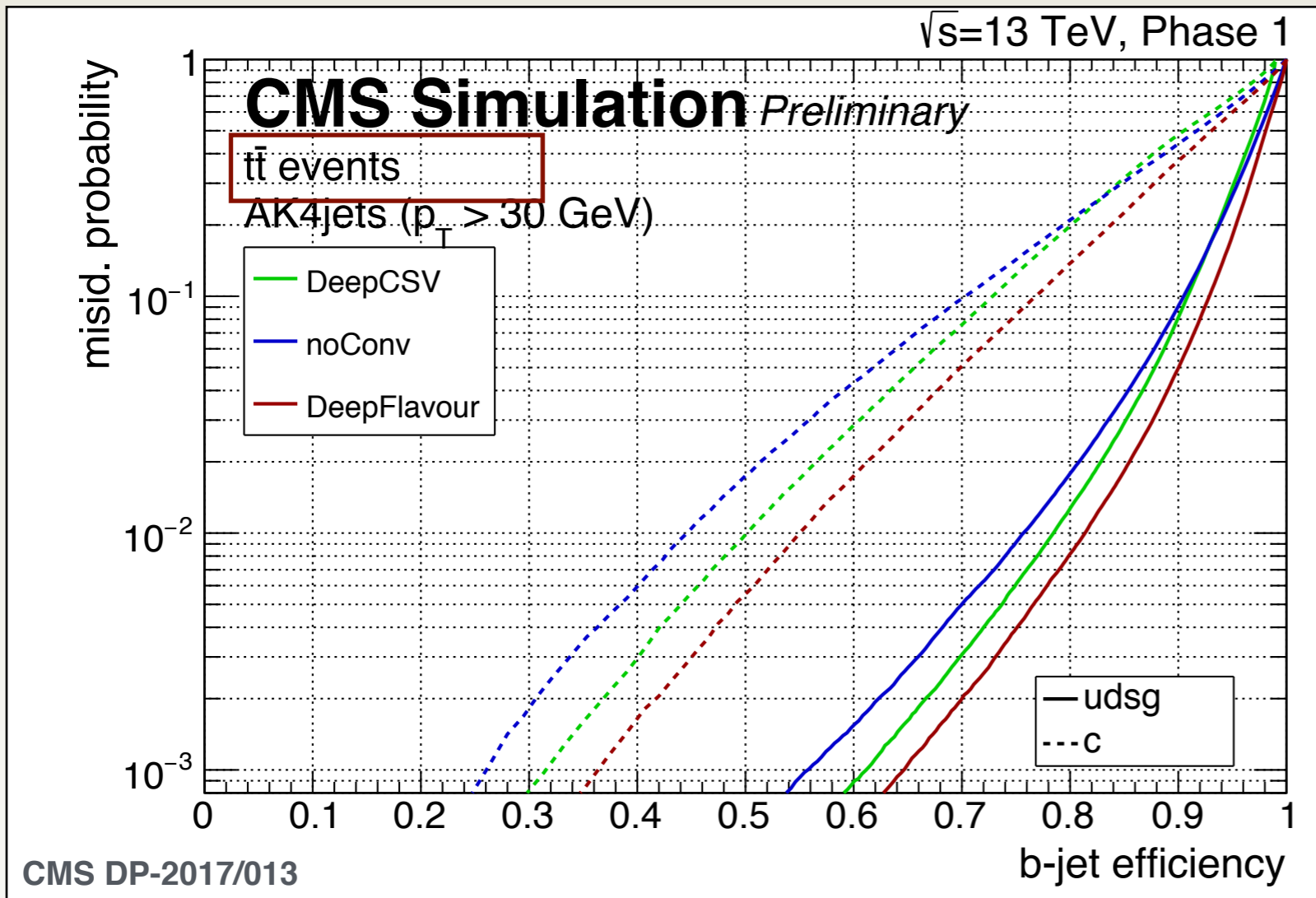
b-Jet Identification: Momentum Dependence



- Typical performance decrease with increasing jet p_T
 - misidentification probability increases
 - b-jet efficiency decreases
- In general, tracks are closer together, more merged tracks etc.
 - Challenging for b tagging
 - Less tracks pass selection criteria
- But: interesting regime for some searches with high p_T b-jets

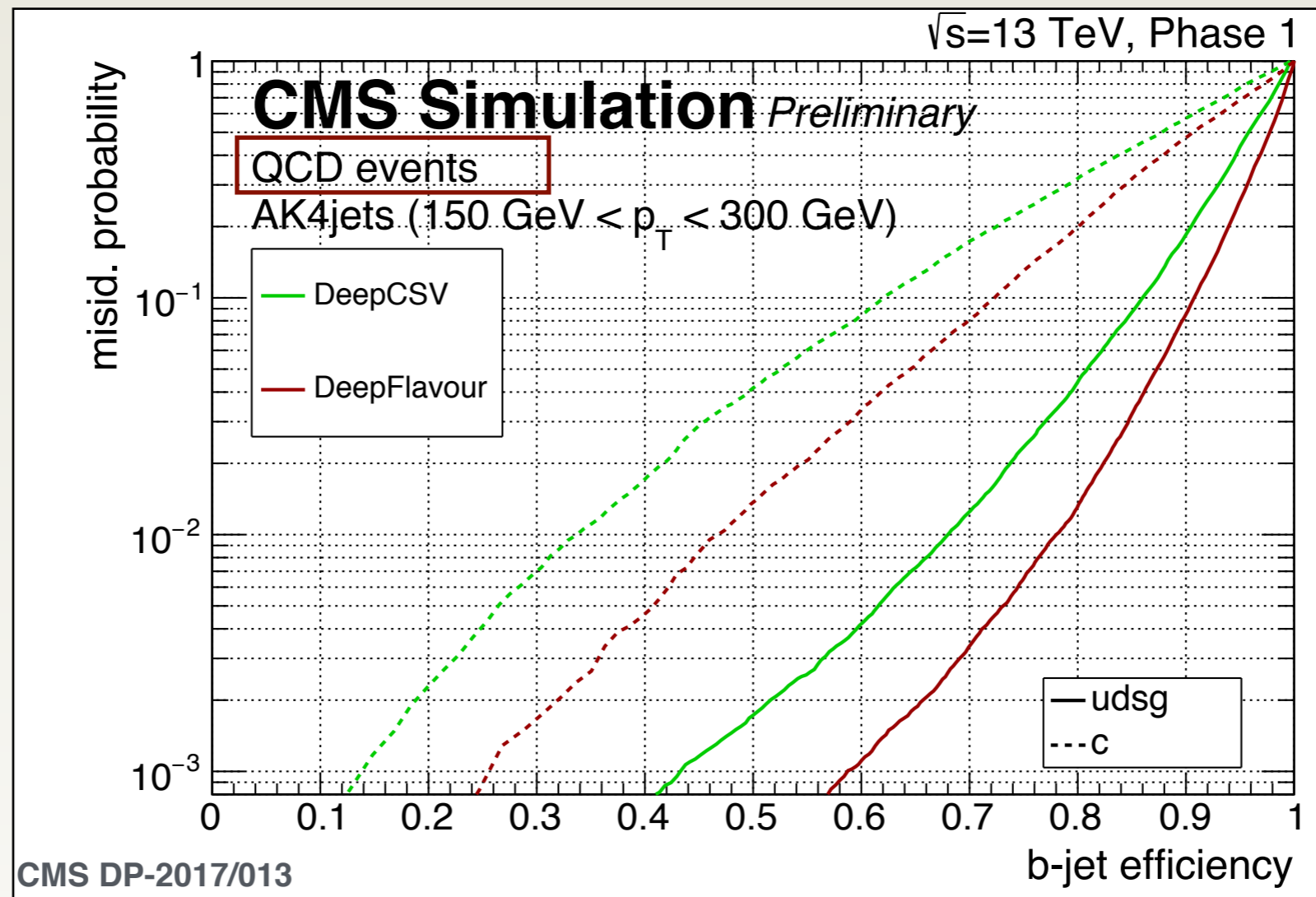


DeepFlavour Jet p_T Dependence



- Top-quark pair events: most jets below 200 GeV
 - Gain from DeepFlavour for moderate b-jet p_T
- What about high p_T jets?

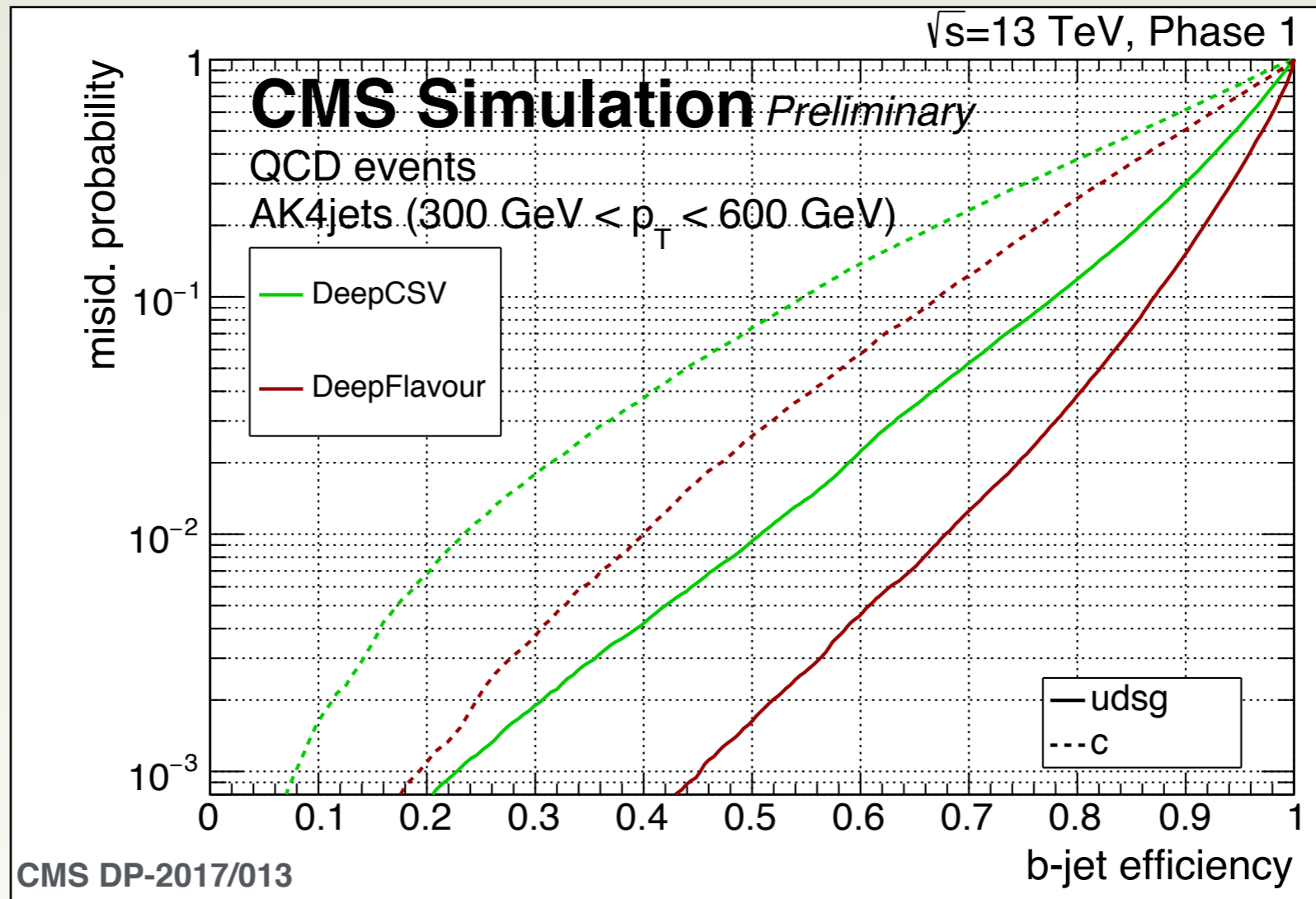
DeepFlavour Jet p_T Dependence



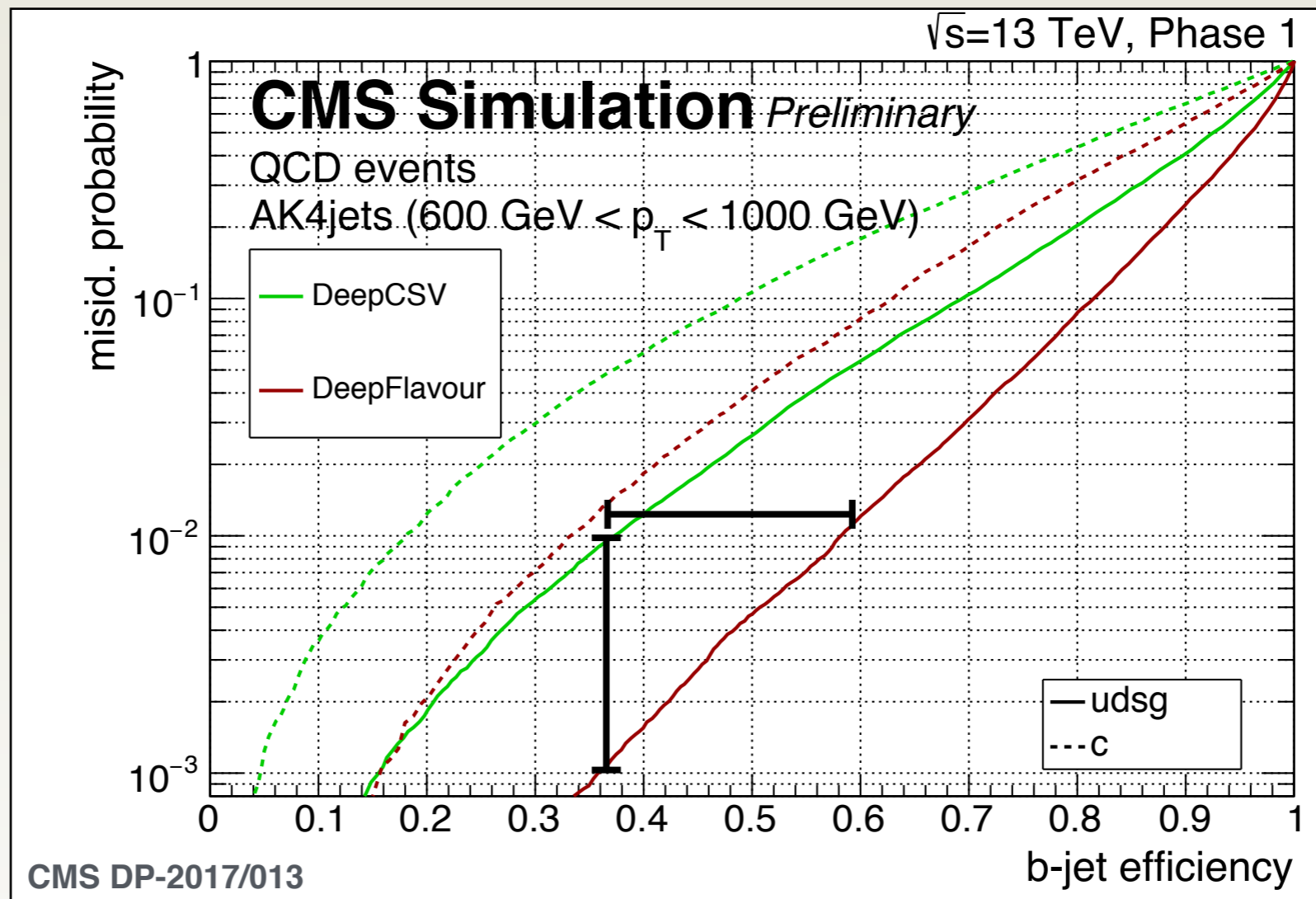
- Use QCD events to estimate the performance at higher jet p_T *
- Already significantly better performance for 150 - 300 GeV

* Different p_T , η range and process - plots not directly comparable to previous slides

DeepFlavour Jet p_T Dependence



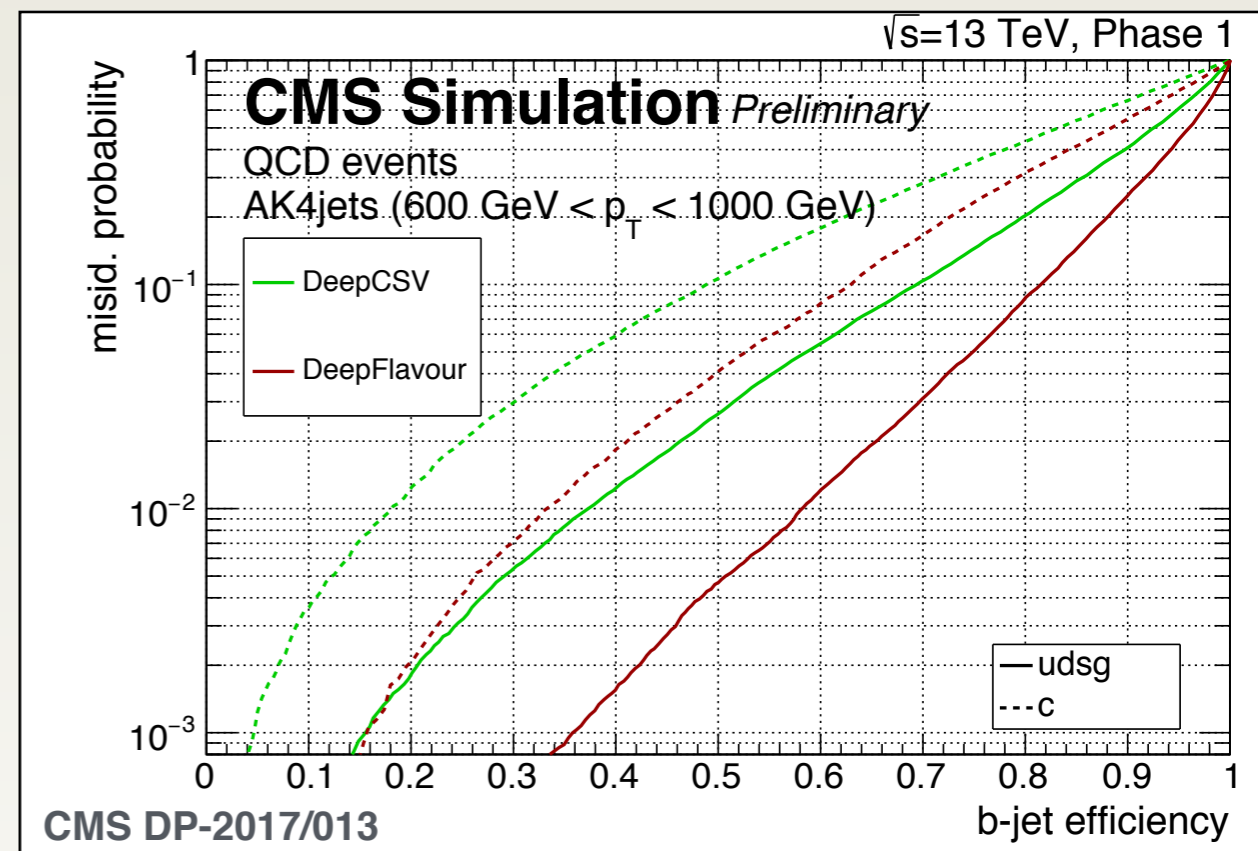
DeepFlavour Jet p_T Dependence



For high p_T

- Up to one order of magnitude better background rejection at same efficiency
- 70% more efficient at same 1% misid. probability (b vs. l)
- Can be a significant gain for analyses with (many) high p_T b-jets.
 - E.g. for identifying 4 out of 4 b jets in this p_T range factor 8 more selected events w.r.t DeepCSV

- Jet-flavour identification can profit a lot from machine-learning techniques
 - DeepCSV shows better performance than other taggers using the same input variables
- Particle-flow candidate based 'DeepFlavour' tagger (under development) with convolutional layers shows outstanding performance



- Considering (almost) all particles in the jet opens possibilities to extend the jet identification to even more categories (taus, leptonic B decays, ...)
- Much more internal studies currently ongoing than public - stay tuned